(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number
WO 02/01950 A2

(51) International Patent Classification[7]: A01K 67/027

(21) International Application Number: PCT/US01/20795

(22) International Filing Date: 29 June 2001 (29.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/215,178	29 June 2000 (29.06.2000)	US
60/215,179	29 June 2000 (29.06.2000)	US
60/215,366	29 June 2000 (29.06.2000)	US
60/215,402	29 June 2000 (29.06.2000)	US
60/215,404	29 June 2000 (29.06.2000)	US
60/215,466	29 June 2000 (29.06.2000)	US
60/215,467	29 June 2000 (29.06.2000)	US
60/221,667	27 July 2000 (27.07.2000)	US
60/244,083	26 October 2000 (26.10.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/215,404 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,366 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,467 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,178 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,179 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,466 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,402 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/244,083 (CON)
Filed on	26 October 2000 (26.10.2000)
US	60/221,667 (CON)
Filed on	27 July 2000 (27.07.2000)

(71) Applicant *(for all designated States except US)*: **DELTAGEN, INC.** [US/US]; 1003 Hamilton Avenue, Menlo Park, CA 94025 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LEVITEN, Michael, W.** [US/US]; 3166 Bryant Street, Palo Alto, CA 94306 (US). **BRENNAN, Thomas, J.** [US/US]; 325 Rockwood Drive, South San Francisco, CA 94080 (US). **GUENTHER, Catherine** [US/US]; 731 Chestnut Street #109, San Carlos, CA 94070 (US). **KLEIN, Robert** [US/US]; 1044 Webster Street, Palo Alto, CA 94301 (US). **MATTHEWS, William** [US/US]; 560 Summit Springs Road, Woodside, CA 94062 (US). **MOORE, Mark** [US/US]; 880 Blandford Avenue, Redwood City, CA 94062 (US).

(74) Agents: **LAPIZ, Mariette, A.** et al.; Deltagen, Inc., 1003 Hamilton Avenue, Menlo PArk, CA 94025 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/01950 A2

(54) Title: TRANSGENIC MICE CONTAINING TARGETED GENE DISRUPTIONS

(57) Abstract: The present invention relates to transgenic animals, as well as compositions and methods relating to the characterization of gene function. Specifically, the present invention provides transgenic mice comprising mutations in a target gene. Such transgenic mice are useful as models for disease and for identifying agents that modulate gene expression and gene function, and as potential treatments for various disease states and disease conditions.

TRANSGENIC MICE CONTAINING TARGETED GENE DISRUPTIONS

Field of the Invention

The present invention relates to transgenic animals, compositions and methods relating to the characterization of gene function.

Background of the Invention

Experimental animal models are important tools for understanding the role of genes. More particularly, the ability to develop animals with specific genes altered or inactivated has been invaluable to the study of gene function, and has lead to unexpected discoveries of a gene and/or mechanisms responsible for disease with similar manifestations in humans. These genetically engineered animals are also useful for identifying and testing therapeutic treatments for a variety of diseases and disorders.

The identification of the function of numerous genes has been useful in ascertaining the roles of these genes in disease. Because of the high level of homology between humans and mice, for example, it is possible to define the function of individual human genes by making targeted germline mutations in selected genes in the animal. The phenotype of the resulting mutant animal can be used to help define the phenotype in humans.

Several interesting genes have recently been discovered belonging to various families encoding G-protein coupled receptors (GPCRs), chordin, nuclear hormone receptors, bone morphogenic proteins, and aquaporin proteins. Identifying the roles of these genes and their expression products may permit the definition of disease pathways and the identification of diagnostically and therapeutically useful targets.

GPCRs have been characterized as having seven putative transmembrane domains (designated TM1, TM2, TM3, TM4, TM5, TM6, and TM7) which are believed to represent transmembrane α-helices connected by extracellular or cytoplasmic loops. Most G-protein coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structure. G-protein coupled receptors can be intracellularly coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels and transporters. Different G-protein α-subunits preferentially stimulate particular effectors to modulate various biological functions in a cell.

Anaphylatoxins, in particular C3a and C5a, have various biological activities which suggest a role as mediators of inflammatory reactions: they cause contraction of smooth muscle, histamine release, increase in capillary permeability, adhesion of leukocytes to vascular endothelium, leukocyte chemotaxis, and aggregation of platelets and leukocytes (*see* Vogt, *Complement* 3:177-88 (1986)). Most of these effects are supported by the cooperation of other mediators, in particular arachidonic acid derivatives which may be produced by anaphylatoxin-stimulated cells, *e.g.* leukocytes or endothelium. In vivo effects of the complement peptides depend very much on the site of their

generation:. intravascular release in the general circulation leads to adverse symptoms such as adult respiratory distress syndrome and shock lung, mainly due to leukocyte activation, aggregation and their accumulation in lung vessels. Intravascular release may be induced by certain drugs, and by contact of blood with the surfaces of bypass or dialysis apparatus. Activation of the complement system has been reported in a variety of inflammatory diseases and neurodegenerative processes of the CNS. Recent evidence indicates that complement proteins and receptors are synthesized on or by glial cells and, surprisingly, neurons. Among these proteins are the receptors for the chemotactic and anaphylactic peptides, C5a and C3a, which are the most-potent mediators of complement inflammatory functions. The functions of glial-cell C3a and C5a receptors (C3aR and C5aR) appear to be similar to immune-cell C3aRs and C5aRs. However, little is known about the roles these receptors might have on neurons. (*see* Nataf, *et al.*, *Trends. Neurosci.* 22:397-402(1999)).

Recently, a cDNA (clone AZ3B) was isolated from a differentiated HL-6 0 cell cDNA library with a probe derived from the N-formyl peptide receptor gene (Roglic *et al.*, *Biochem. Biophys. Acta.* 1305(1-2):39-43 (1996); accession no. U28488; GI:1199577). The 1.97-kb cDNA encoded a putative G protein-coupled receptor with 482 amino acids. In addition to the predicted 7 transmembrane domains common to all GPCRs, the protein contained a large extracellular loop of approximately 172 amino acids between the fourth and the fifth transmembrane domains, a feature unique among the hundreds of GPCRs identified to date. High sequence homology existed between the protein and a number of chemoattractant receptors in the amino-terminal 170 residues and the carboxyl-terminal 150 residues, such as human C3a anaphylatoxin receptor (C3AR)(Crass *et al.*, *Eur. J. Immunol.* 26(8):1944-50(1996)) as well as murine complement C3AR gene (accession no. AF053757; GI:3170513). Northern and flow cytometric analyses suggested that the message and protein are widely expressed in several differentiated hematopoietic cell lines, in the lung, placenta, heart, and endothelial cells. Roglic *et al.* postulated that the protein defines a distinct group of receptors within the GPCR superfamily. Tornetta, *et al.* reported a full-length C3a cDNA of 3.3-kb encoding an open reading frame of 477 amino acids (*J. Immunol.* 158:5277-82 (1997); accession no. U77461; GI 2130534) The predicted amino acid contained four predicted N-linked glycosylation sites and was 65% identical to the 482 amino acids comprising the coding region of the human C3a receptor.

One important subfamily of the GPCRs is the serotonin neurotransmitter receptor GPCRs. Serotonin (*aka* 5-HT) is an important neuromodulator and local hormone in the CNS and intestine that mediates a wide range of physiological functions and pathophysiological pathways by activating multiple receptors. About 90% of serotonin in the body is present in chromaffin cells in the wall of the intestine with a small amount in the myenteric plexus. Serotonin is also found in blood where it is present in high concentration in platelets. In the CNS, serotonin neurons originate primarily in the raphe nuclei of the brainstem and project to most areas of the brain. It is synthesized from the amino acid tryptophan, via 5-hydroxytryptophan. In the periphery, 5-HT contracts a number of smooth

muscles including large blood vessels, intestine and uterus, and also induces endothelium-dependent vasodilation through the formation of NO. Serotonin stimulates sensory nerve endings and is a mediator of peristalsis and may be involved in platelet aggregation and hemostasis. It may also have an inflammatory mediator and involvement in microvascular control. In the CNS, serotonin is thought to be involved in a wide range of functions including control of appetite, mood, anxiety, hallucinations, sleep, vomiting and pain perception. Serotonin receptor ligands have found clinical use in the treatment of depression, migraine and post-operative vomiting, and there is a strong potential for their use in other conditions.

Molecular biology studies as well as pharmacological studies have revealed the existence of a large number of subtypes of 5-HT receptors. One serotonin receptor is the 5-HT-5 receptor. In the CNS, the 5-HT-5 receptor mRNA is found in the cerebral cortex, hippocampus, habenula, olfactory bulb and granular layer of the cerebellum. The cDNA (1686 bp) encoding the 5-HT5 (A subtype) receptor was isolated from a mouse brain library (EMBO J. 11(13), 4779-86 (1992)). This sequence has been submitted to GenBank (GI or NID number: 49758; Accession number: Z18278). Amino acid sequence comparisons revealed that this 5-HT5A receptor was a distant relative of all previously identified 5-HT receptors. Its pharmacological profile resembled that of the 5-HT-1D receptor, which is a 5-HT receptor subtype which has been shown to inhibit adenylate cyclase and which is predominantly expressed in basal ganglia. However, unlike 5-HT-1D receptors, the 5-HT5A receptor did not inhibit adenylate cyclase and its mRNA was not found in basal ganglia. The coding sequence for the 5-HT5A gene is believed to comprise bases 509-1582.

Over the past 15 years, nearly 350 therapeutic agents targeting 7 transmembrane receptors have been successfully introduced onto the market. As these receptors have an established, proven history as therapeutic targets, a clear need exists for identification and characterization of GPCRs which can play a role in preventing, ameliorating or correcting dysfunctions or diseases.

Generally speaking, growth factors are proteins that bind to receptors on the cell surface, with the primary result of activating cellular proliferation and/or differentiation. Many growth factors are quite versatile, stimulating cellular division in numerous different cell types; while others are specific to a particular cell-type. More specifically, growth factors are substances, such as polypeptide hormones, which affect the growth of defined populations of animal cells *in vivo* or *in vitro*, but which are not nutrient substances. Proteins involved in the growth and differentiation of tissues may promote or inhibit growth, and promote or inhibit differentiation, and thus the general term "growth factor" includes cytokines, trophic factors, and their inhibitors. Among growth, or neurotrophic, factors presently known are the transforming growth factors (TGF-alpha, TGF-beta, TGF-gamma). Transforming growth factor-beta appears to elicit a variety of responses in many different cell types.

Among TGF-beta members are the bone morphogenetic proteins (BMP). The BMPs have been indicated as useful in wound healing, tissue repair, and to induce cartilage and/or bone growth.

BMPs have potent effects during embryogenesis. One particular member, BMP-4, has been shown to have potent ventralizing effects in *Xenopus* embryos, leading to the differentiation of blood and mesenchyme and inhibiting the formation of dorsal tissues such as notochord, muscle, and nervous system. (*See, e.g.*, Jones *et al.*, *Development* 115:639-647 (1991)). BMP-4 is expressed ventrally in the *Xenopus* embryo. The expression of BMP-4 is increased by ventralizing treatments such as irradiation with ultraviolet light. Inhibitors of ventralizing BMPs are believed to have dorsalizing effects on tissue differentiation. One such growth factor inhibitor is chordin.

Chordin (CHRD) is a key developmental growth factor inhibitor that dorsalizes early vertebrate embryonic tissues by binding to ventralizing TGF-beta-like BMPs and sequestering them in latent complexes. Specifically, CHRD is a secreted BMP antagonist. Sequence databases for mammalian chordin sequences searched and ESTs encoding the C-terminal regions of human CHRD and mouse CHRD were identified. It was determined that the full-length cDNA sequence for mouse CHRD predicted a 948-amino acid protein containing 4 internal cysteine-rich repeats and 4 potential N-glycosylation sites. (*See, e.g.*, Pappano *et al.* (*Genomics* 52: 236-239 (1998)). Northern blot analysis detected high levels of a 3.7-kb CHRD transcript in 7-day postcoitum mouse embryos. CHRD transcript levels decreased at later developmental stages and in adult tissues, implying a major role for chordin during gastrulation of the mammalian embryo. Additionally, both murine and human chordin genes were shown to be expressed at readily detectable levels in several fetal and adult tissues, most notably liver and cerebellum, suggesting additional roles in organogenesis and homeostasis. Dot blot analysis of mRNA expression in human adult and fetal tissues showed that CHRD is expressed at the highest level in liver. A partial human CHRD cDNA sequence was deposited in GenBank (GI or NID number: 3822217; Accession number: AF076612).

A human CHRD cDNA, which encodes a 954-amino acid protein that has 86% sequence identity with the mouse protein was cloned and sequenced. (Smith *et al.*, *Hum. Genet.* 105:104-111 (1999)). It was also determined that the CHRD gene has 23 exons spanning 11.5 kb. The human CHRD gene was localized to 3q27 by radiation hybrid mapping and the mouse CHRD gene to proximal chromosome 16 using interspecific backcrosses. (*See*, Pappano *et al.*, *supra*). The genomic organization of the human chordin gene was determined and shown to map within a gene cluster at 3q27 containing THPO (thrombopoietin), CLCN2 (a voltage-gated chloride-channel gene) and EIF4G1 (a eukaryotic translation-initiation-factor-gamma gene). (*See*, Smith *et al.*, *supra*). The CHRD and THPO genes are very close neighbors (within a single cosmid clone). The two genes occupy a total of 21 kb and are transcribed from opposite strands using promoters separated by less than 2 kb. The CHRD gene and the chordin-regulating GSC (goosecoid) gene were found to be potential candidate genes for Cornelia de Lange syndrome (CDLS), a developmental malformation syndrome mapped to 3q, which is primarily characterized by mental handicap, growth retardation, distinctive facial features and limb-reduction defects. CDLS patients typically occur as sporadic cases,

but several reports have suggested dominant inheritance. The candidacy of the CHRD and GSC genes was supported by several lines of evidence including: prior evidence for a CDLS gene at 3q26.3-q27; a report suggesting a significant association between CDLS and thrombocytopenia; suspected genetic heterogeneity in CDLS; location of the GSC gene in close proximity to a 14q32 breakpoint detected in a CDLS patient with a balanced de novo translocation; known regulation of chordin expression by goosecoid; and the pattern of embryonic expression of the mouse GSC gene. Mutation screening failed, however, to identify CDLS patient-specific mutations in CHRD or GSC.

Mouse CHRD mRNA was first expressed in the anterior primitive streak and then in the node and axial mesoendoderm that derives from it, suggesting that CHRD may play a role in patterning of the early embryo. Targeted inactivation of CHRD resulted in stillborn animals having normal early development and neural induction, but displaying later defects in inner and outer ear development and abnormalities in pharyngeal and cardiovascular organization.

It was demonstrated that at mid-gastrula, expression of noggin overlaps that of CHRD. (*See, e.g.*, Bachiller *et al.* (*Nature* 403: 658-661 (2000)). Noggin mutants underwent normal gastrulation and anterior central nervous system patterning, although at later stages a number of abnormalities were observed in the posterior spinal cord and somites. Intercrosses between mice heterozygous for noggin and chordin mutations did not produce double-homozygous mutants among the neonates. Two chordin/noggin double-null embryos were found among animals dissected close to term. Both were undergoing resorption, but clearly had holoprosencephaly, with a single nasal pit, a cyclopic eye, and agnathia. These malformations, not observed in either mutant on its own resembled embryos lacking sonic hedgehog (shh). At embryonic day 12.5, double-mutant embryos were recovered with more severe phenotypes resembling aprosencephaly. In double-mutant embryos dissected at embryonic day 8.5, forebrain reduction was clearly evident. This data demonstrated that chordin and noggin were not necessary for establishing the anterior visceral endoderm, but are required for subsequent elaboration of anterior pattern. Mesodermal development was also affected, indicated by the lack of shh. It was suggested that the BMP antagonists chordin and noggin compensate for each other during early mouse development. When both gene products were removed, antero-posterior, dorso-ventral, and left-right patterning were all affected.

Recently, a 3254 bp sequence for a mouse CHRD cDNA was reported (Genbank accession no. AF096276; GI 4406185). The coding region was believed to span nucleotides 1-2847, encoding a 948 amino acid polypeptide. Given the importance of growth factor inhibitors, a clear need exists for identification and characterization of growth factor inhibitors which can play a role in preventing, ameliorating or correcting dysfunctions or diseases.

The nuclear receptor superfamily constitutes a diverse group of ligand-activated transcriptional factors that share a common modular structure. This superfamily includes receptors for steroid hormones, thyroid hormone, retinoids and vitamin D as well as a large number of receptors, referred

to as orphan receptors, for which no ligand has been identified. The superfamily can be divided into many subfamilies, each containing several closely related genes. In addition to RORα and RORβ, a third member of the ROR subfamily, RORγ, has been cloned (Hirose *et al.*, *Biochem. Biophys. Res. Commun.* 205:1976-1983 (1994); Medvedev *et al.*, *Gene* 181:199-206 (1996)); Medvedev *et al.*, *Genomics*, 46:93-102 (1997)). The amino acid sequence of RORγ is highly conserved between species; the amino acid sequence between mouse and human RORγ exhibit 88% identity (Hirose *et al.*, (1994); Medvedev *et al.*, (1996); Medvedev *et al.*, (1997)).

Most nuclear receptors contain five major domains: an A/B domain at the amino-terminus, a highly conserved DNA-binding (C) domain containing two zinc finger motifs, a hinge region (D), a ligand-binding domain (E), and an F domain at the carboxyl-terminus. The nuclear receptors bind as monomers or as homo- or heterodimers to response elements composed of a single core motif containing the consensus sequence PuGGTCA or to direct, palindromic, or inverted repeats of the core motif spaced by one or more nucleotides. Roles for nuclear receptors have been demonstrated in the control of embryonic development, cell proliferation and differentiation. In addition, genetic alterations in the expression and structure of these receptors have been implicated in various disease processes (Hughes *et al.*, *Science* 242:1702-1705 (1988); DeLucas, *FASEB J.* 5:2924-2933 (1991); Chomienne *et al.*, *Leukemia* 4:802-807 (1991)).

Recently, a cDNA (mRORγ orphan nuclear receptor) mRNA was isolated from a mouse muscle cDNA library (Medvedev, *et al.*, *Gene* 181199-206 (1996); accession no. U4350; GI: 1155340). The 2066 bp cDNA encoded a protein of 516 amino acid residues, similar to the human 560 amino acid ROR (*see* Hirose, *et al.*, *Biochem. Biophys. Res. Commun.* 205:1976-83 (1994)). Furthermore, high sequence homology existed between the protein and human ROR gamma, with an overall identity of 88%. Analysis of the ROR gamma-response element using in vitro synthesized ROR gamma revealed that it binds as a monomer to response elements composed of a single core motif GGTCA preceded by a 6 bp AT-rich sequence. Northern blot analysis using RNA from different tissues showed that mROR gamma was found to be highly expressed in skeletal muscle, liver and kidney.

Thus, while important regulatory functions have been suggested generally for members of the ROR subfamily (Austin et al., *Cell Growth & Differentiation*, 9:267-276 (1998)), the particular role played by RORγ in development and differentiation has yet to be elucidated.

Among TGF-beta members are the bone morphogenetic proteins (BMP). The BMPs have been indicated as useful in wound healing, tissue repair, and to induce cartilage and/or bone growth. BMPs have potent effects during embryogenesis. One member, BMP-4, has been shown to have potent ventralizing effects in *Xenopus* embryos, leading to the differentiation of blood and mesenchyme and inhibiting the formation of dorsal tissues such as notochord, muscle, and nervous system (*see, e.g.*, Jones *et al.*, *Development* 115:639-647 (1991)). .BMP-4 is expressed ventrally in

the *Xenopus* embryo and its expression is increased by ventralizing treatments such as irradiation with ultraviolet light.

During animal development, cells exchange signals with their neighbors that allow specific groups of cells to produce organized, differentiated structures such as limbs, segments or other body parts. Understanding these processes might allow one to manipulate them to regenerate damaged body parts or to diagnose birth defects. The twisted gastrulation protein (TSG) is one of five known secreted proteins required to pattern the dorsal part of the early *Drosophila* embryo. Unlike the decapentaplegic (DPP) protein that is required to pattern the entire dorsal half of the embryo, TSG is needed only to specify the fate of the dorsal midline cells. When expressed in a ventral stripe of cells, TSG protein can diffuse to the dorsal-most cells and can rescue the dorsal midline cells in tsg mutant embryos. TSG functions in a permissive rather than instructive role to elaborate cell fates along the dorsal midline after peak levels of DPP activity have 'primed' cells to respond to TSG.

The serine proteases are secretory proteins which contain N-terminal signal peptides that serve to export the immature protein across the endoplasmic reticulum and are then cleaved (von Heijne, *Nucl. Acid. Res.* 14:5683-90(1986)). Differences in these signal sequences provide one means of distinguishing individual enzymes. Some serine proteases, particularly the digestive enzymes, exist as inactive precursors or preproenzymes, and contain a leader or activation peptide sequence 3' of the signal peptide. Other features that distinguish various serine proteases are the presence or absence of N-linked glycosylation sites that provide membrane anchors, the number and distribution of cysteine residues that determine the secondary structure of the serine protease, and the sequence of a substrate binding sites such as S'. The S' substrate binding region is defined by residues extending from approximately +17 to +29 relative to the N-terminal I (+1). Differences in this region of the molecule are believed to determine enzyme substrate specificities.

Proteases participate in a variety of developmental and metabolic processes (Stroud, *Sci. Am.* 231:74-88 (1974); Neurath, *Science* 224:350-357 (1984)). Molecular defects that alter enzyme function often lead to serious human diseases, such as bleeding, thrombosis and atherosclerosis. Hepsin is a novel serine protease of the trypsin family containing a transmembrane domain near its amino-terminus (Kurachi *et al.*, *Methods in Enzymology*, 244:100-114 (1994)). This structural feature distinguishes hepsin from most other serine proteases. In addition to blood coagulation, hepsin is reported to be critical for cell growth. Transgenic mice having a disrupted hepsin gene exhibit elevated blood serum alkaline phosphatase levels (U.S. patent no. 5,981,830; Wu *et al.*, *J. Clin. Invest.* 101:321-326 (1998)). No obvious abnormalities were found in major organs in hepsin -/- mice on histologic examination. The results indicated that hepsin was not essential for embryonic development and normal hemostasis.

In the fruitfly *Drosophila melanogaster*, another transmembrane serine protease, Stubble-stubbloid, has been reported that has similar structural features as that of hepsin. Stubble-stubbloid

protein plays an important role in epithelial morphogenesis and development in *Drosophila*. Defects in Stubble-stubbloid gene cause malformation of legs, wings and bristles (Appel *et al.*, *Proc. Natl. Acad. Sci. USA* 90:4937-4941 (1993)).

Airway trypsin-like protease is an enzyme isolated from the sputum of patients with chronic airway diseases (Yamaoka *et al.*, *J. Biol. Chem.* 273(19):11895-901(1998); accession NP 004253; GI:4758508). An entire 1517-base pair sequence of cDNA was obtained with an open reading frame encoding a polypeptide with 418-amino acid residues. The polypeptide consisted of a 232-residue catalytic region and a 186-residue noncatalytic region with a hydrophobic putative transmembrane domain near the NH2 terminus. The polypeptide was suggested to be a type II integral membrane protein in which the COOH-terminal catalytic region is extracellular. This protein was thought to be synthesized as a membrane-bound precursor and to mature to a soluble and active protease by limited proteolysis. It showed 29-38% identity in the sequence of the catalytic region with human hepsin, enteropeptidase, acrosin and mast cell tryptase. The noncatalytic region had little similarity to other known proteins. In Northern blot analysis a transcript of 1.9 kilobases was detectable most prominently in the trachea among 17 human tissues examined. Recently, an mouse EST has been identified from a Knowles Solter embryo cDNA library (accession no. AA415658; GI:2075923) bearing sequence similarity to human airway trypsin-like protease and to human hepsin.

The aquaporin family is a group of selective water channels, which is a part of the larger family of major intrinsic membrane proteins (MIP). To date, eight different aquaporins have been cloned. Aquaporins-1 through –4 have been identified within the kidney and play a role in the water reabsorption of the glomerular filtrate along the nephron. Aquaporin 0 is the major (60%) intrinsic protein of lens fiber cells of the eye. Mutations in this gene are associated with cataract formation in mice. Aquaporin 1, also called CHIP-28, exists in the membrane as a homotetramer, and is present in red blood cells, the choroid plexus, the proximal tubule and descending limb of the loop of Henle in the kidney as well as in many other sites (Heymann *et al. JSB* 121(2), 191-206 (1998)). Surprisingly, no pathological consequence is known in patients lacking a functional AQP1 gene. Aquaporin 2, also called WCH-CD, is the water channel of the principal cell of the cortical and medullary collecting duct, and is located in cytoplasmic vesicles unless arginine vasopressin is acting, when it is translocated to the apical membrane by synaptobrevins or vesicle associated membrane protein 2 (VAMP2). Lack of functional AQP2 gene leads to a rare form of nephrogenic diabetes insipidus. AQP3 and AQP4 are localized to the basolateral membrane of principal cells of the collecting ducts, AQP3 being located predominantly in the cortical and outer medulla, whereas AQP4 predominates in the inner medullary collecting duct. The clinical importance of these two water channels in the renal water conservation has yet to be determined. Recently, a transgenic mouse deficient in AQP4 was generated (Ma *et al.*, *J. Clin. Invest.* 100:957-962 (1997)). Interestingly, the phenotype abnormalities were very subtle, with a small defect in maximal urinary concentration, consistent with the

predominant localization of AQP4 in the lower portion of the IMCD. Aquaporin-5 apparently is coupled to fluid secretion in exocrine tissues. Although the exact function of aquaporin-6 is not known due to its uncertain localization, its restricted presence in the kidney may suggest a potential role in water transport. Aquaporin-7 appears to play a role in the cryopreservation of the sperm whereas aquaporin-8 is thought to be responsible for the secretion of pancreatic juice (*reviewed in* Dibas, *et al., Proc. Soc. Exp. Biol. Med.* 219:183-99(1998)). In general, very little work has examined the ontogeny of these proteins, except in the rat, and virtually nothing is known of the expression of these genes in pregnancy or in any disorder of fluid balance in the mother or fetus.

Recently, a mouse aquaporin-8 water channel cDNA was isolated (accession no. AF018952; GI:2353796). The 1447 bp cDNA, (designated AQP8) was found to have a 783 bp open reading frame. The encoded 261 amino acid hydrophobic protein contains the conserved NPA motifs found in MIP family members (Ma, *et al., Biochem. Biophys. Res. Commun.* 240:324-8(1997)).

Summary of the Invention

The present invention generally relates to transgenic animals, as well as to compositions and methods relating to the characterization of gene function.

The present invention provides transgenic cells comprising a disruption in a targeted gene. The transgenic cells of the present invention are comprised of any cells capable of undergoing homologous recombination. Preferably, the cells of the present invention are stem cells and more preferably, embryonic stem (ES) cells, and most preferably, murine ES cells. According to one embodiment, the transgenic cells are produced by introducing a targeting construct into a stem cell to produce a homologous recombinant, resulting in a mutation of the targeted gene. In another embodiment, the transgenic cells are derived from the transgenic animals described below. The cells derived from the transgenic animals includes cells that are isolated or present in a tissue or organ, and any cell lines or any progeny thereof.

The present invention also provides a targeting construct and methods of producing the targeting construct that when introduced into stem cells produces a homologous recombinant. In one embodiment, the targeting construct of the present invention comprises first and second polynucleotide sequences that are homologous to the targeted gene. The targeting construct also comprises a polynucleotide sequence that encodes a selectable marker that is preferably positioned between the two different homologous polynucleotide sequences in the construct. The targeting construct may also comprise other regulatory elements that may enhance homologous recombination.

The present invention further provides non-human transgenic animals and methods of producing such non-human transgenic animals comprising a disruption in a target gene. The transgenic animals of the present invention include transgenic animals that are heterozygous and homozygous for a mutation in the target gene. In one aspect, the transgenic animals of the present

invention are defective in the function of the target gene. In another aspect, the transgenic animals of the present invention comprise a phenotype associated with having a mutation in a target gene.

The present invention also provides methods of identifying agents capable of affecting a phenotype associated with a disruption of a target gene in a transgenic animal. For example, a putative agent is administered to the transgenic animal and a response of the transgenic animal to the putative agent is measured and compared to the response of a "normal" or wild type mouse, or alternatively compared to a transgenic animal control (without agent administration). The invention further provides agents identified according to such methods. The present invention also provides methods of identifying agents useful as therapeutic agents for treating conditions associated with a disruption of the target gene.

The present invention further provides a method of identifying agents having an effect on target gene expression or function. The method includes administering an effective amount of the agent to a transgenic animal, preferably a mouse. The method includes measuring a response of the transgenic animal, for example, to the agent, and comparing the response of the transgenic animal to a control animal, which may be, for example, a wild-type animal or alternatively, a transgenic animal control. Compounds that may have an effect on target gene expression or function may also be screened against cells in cell-based assays, for example, to identify such compounds.

The invention also provides cell lines comprising nucleic acid sequences of a target gene. Such cell lines may be capable of expressing such sequences by virtue of operable linkage to a promoter functional in the cell line. Preferably, expression of the target gene sequence is under the control of an inducible promoter. Also provided are methods of identifying agents that interact with the target gene, comprising the steps of contacting the target gene or target protein with an agent and detecting an agent/ target gene or agent/target protein complex. Such complexes can be detected by, for example, measuring expression of an operably linked detectable marker.

The invention further provides methods of treating diseases or conditions associated with a disruption in a target gene, and more particularly, to a disruption in the expression or function of the target gene or the target protein encoded by the target gene. In a preferred embodiment, methods of the present invention involve treating diseases or conditions associated with a disruption in the target gene's expression or function, including administering to a subject in need, a therapeutic agent which effects target gene or target protein expression or function. In accordance with this embodiment, the method comprises administration of a therapeutically effective amount of a natural, synthetic, semi-synthetic, or recombinant target gene, target gene products or fragments thereof as well as natural, synthetic, semi-synthetic or recombinant analogs.

The present invention further provides methods of treating diseases or conditions associated with disrupted targeted gene expression or function, wherein the methods comprise detecting and replacing through gene therapy mutated target genes.

Definitions

The term "gene" refers to (a) a gene containing at least one of the DNA sequences disclosed herein; (b) any DNA sequence that encodes the amino acid sequence encoded by the DNA sequences disclosed herein and/or; (c) any DNA sequence that hybridizes to the complement of the coding sequences disclosed herein. Preferably, the term includes coding as well as noncoding regions, and preferably includes all sequences necessary for normal gene expression including promoters, enhancers and other regulatory sequences.

The terms "polynucleotide" and "nucleic acid molecule" are used interchangeably to refer to polymeric forms of nucleotides of any length. The polynucleotides may contain deoxyribonucleotides, ribonucleotides and/or their analogs. Nucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The term "polynucleotide" includes single-, double-stranded and triple helical molecules.

"Oligonucleotide" refers to polynucleotides of between 5 and about 100 nucleotides of single- or double-stranded DNA. Oligonucleotides are also known as oligomers or oligos and may be isolated from genes, or chemically synthesized by methods known in the art. A "primer" refers to an oligonucleotide, usually single-stranded, that provides a 3'-hydroxyl end for the initiation of enzyme-mediated nucleic acid synthesis. The following are non-limiting embodiments of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A nucleic acid molecule may also comprise modified nucleic acid molecules, such as methylated nucleic acid molecules and nucleic acid molecule analogs. Analogs of purines and pyrimidines are known in the art, and include, but are not limited to, aziridinycytosine, 4-acetylcytosine, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethyl-aminomethyluracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, pseudouracil, 5-pentylnyluracil and 2,6-diaminopurine. The use of uracil as a substitute for thymine in a deoxyribonucleic acid is also considered an analogous form of pyrimidine.

A "fragment" of a polynucleotide is a polynucleotide comprised of at least 9 contiguous nucleotides, preferably at least 15 contiguous nucleotides and more preferably at least 45 nucleotides, of coding or non-coding sequences.

The term "gene targeting" refers to a type of homologous recombination that occurs when a fragment of genomic DNA is introduced into a mammalian cell and that fragment locates and recombines with endogenous homologous sequences.

The term "homologous recombination" refers to the exchange of DNA fragments between two DNA molecules or chromatids at the site of homologous nucleotide sequences.

The term "homologous" as used herein denotes a characteristic of a DNA sequence having at least about 70 percent sequence identity as compared to a reference sequence, typically at least about 85 percent sequence identity, preferably at least about 95 percent sequence identity, and more preferably about 98 percent sequence identity, and most preferably about 100 percent sequence identity as compared to a reference sequence. Homology can be determined using a "BLASTN" algorithm. It is understood that homologous sequences can accommodate insertions, deletions and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align. The reference sequence may be a subset of a larger sequence, such as a portion of a gene or flanking sequence, or a repetitive portion of a chromosome.

The term "target gene" (alternatively referred to as "target gene sequence" or "target DNA sequence" or "target sequence") refers to any nucleic acid molecule or polynucleotide of any gene to be modified by homologous recombination. The target sequence includes an intact gene, an exon or intron, a regulatory sequence or any region between genes. The target gene comprises a portion of a particular gene or genetic locus in the individual's genomic DNA. As provided herein, the target gene of the present invention is an anaphylatoxin C3a receptor gene, a 5-HT5A gene, an RORγ gene, a BMP gene, an airway trypsin-like protease gene, or an aquaporin-8 gene.

An "anaphylatoxin C3a receptor gene" refers to a sequence comprising SEQ ID NO:1 or comprising the sequence identified in Genebank as Accession No. U77461; GI 2130534. In one aspect, the coding sequence of the anaphylatoxin C3a receptor gene comprises SEQ ID NO:1 or comprises the gene identified in Genebank as Accession No. U77461; GI 2130534.

An "5-HT5A gene" refers to a sequence comprising SEQ ID NO:5 or comprising the sequence identified in Genebank as Accession No. Z18278; GI NO: 49758. In one aspect, the coding sequence of the 5-HT5A gene comprises SEQ ID NO:5 or comprises the gene identified in Genebank as Accession No. Z18278; GI NO: 49758.

A "chordin gene" refers to a sequence comprising SEQ ID NO:9 or comprising the sequence identified in Genebank as Accession No. AF096276; GI 4406185. In one aspect, the coding sequence of the chordin gene comprises SEQ ID NO:9 or comprises the gene identified in Genebank as Accession No. AF096276; GI 4406185.

An "RORγ gene" refers to a sequence comprising SEQ ID NO:13 or comprising the sequence identified in Genebank as Accession No. U4350; GI NO: 1155340. In one aspect, the coding sequence of the RORγ gene comprises SEQ ID NO:13 or comprises the gene identified in Genebank as Accession No. U4350; GI NO: 1155340.

A "BMP gene" refers to a sequence comprising SEQ ID NO:17 or comprising the sequence identified in GenBank as Accession No.: AF292033; GI NO: 9837569. In one aspect, the coding

sequence of the BMP gene comprises SEQ ID NO:17 or comprises the gene identified in Genebank as Accession No.: AF292033; GI NO: 9837569.

An "airway trypsin-like protease gene" refers to a sequence comprising SEQ ID NO:21 or comprising the sequence identified in Genebank as Accession No.: AA415658; GI NO: 2075923. In one aspect, the coding sequence of the airway trypsin-like protease gene comprises SEQ ID NO:21 or comprises the airway trypsin-like protease gene identified in Genebank as Accession No.: AA415658; GI NO: 2075923.

An "aquaporin-8 gene" refers to a sequence comprising SEQ ID NO:25 or comprising the sequence identified in Genebank as Accession No.: AF018952; GI NO: 2353796. In one aspect, the coding sequence of the aquaporin-8 gene comprises SEQ ID NO:25 or comprises the aquaporin-8 gene identified in Genebank as Accession No.: AF018952; GI NO: 2353796.

"Disruption" of a target gene occurs when a fragment of genomic DNA locates and recombines with an endogenous homologous sequence. These sequence disruptions or modifications may include insertions, missense, frameshift, deletion, or substitutions, or replacements of DNA sequence, or any combination thereof. Insertions include the insertion of entire genes which may be of animal, plant, prokaryotic, or viral origin. Disruption, for example, can alter or replace a promoter, enhancer, or splice site of a target gene, and can alter the normal gene product by inhibiting its production partially or completely or by enhancing the normal gene product's activity.

The term, "transgenic cell", refers to a cell containing within its genome a target gene that has been disrupted, modified, altered, or replaced completely or partially by the method of gene targeting.

The term "transgenic animal" refers to an animal that contains within its genome a specific gene that has been disrupted by the method of gene targeting. The transgenic animal includes both the heterozygote animal (*i.e.*, one defective allele and one wild-type allele) and the homozygous animal (*i.e.*, two defective alleles).

As used herein, the terms "selectable marker" or "positive selection marker" refers to a gene encoding a product that enables only the cells that carry the gene to survive and/or grow under certain conditions. For example, plant and animal cells that express the introduced neomycin resistance (Neo^r) gene are resistant to the compound G418. Cells that do not carry the Neo^r gene marker are killed by G418. Other positive selection markers will be known to those of skill in the art.

A "host cell" includes an individual cell or cell culture which can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent due to natural, accidental, or deliberate mutation. A host cell includes cells transfected with the constructs of the present invention.

The term "modulates" as used herein refers to the inhibition, reduction, increase or enhancement of a target gene or target protein function, expression, activity, or alternatively a phenotype associated with a disruption in a target gene.

The term "ameliorates" refers to a decreasing, reducing or eliminating of a condition, disease, disorder, or phenotype, including an abnormality or symptom associated with a disruption in a target gene.

The term "abnormality" refers to any disease, disorder, condition, or phenotype in which a disruption of a target gene is implicated, including pathological conditions.

Brief Description of the Drawings

Figure 1 shows the polynucleotide sequence for an orphan anaphylatoxin C3a receptor (SEQ ID NO:1). Figure 1 also shows the amino acid sequence for the anaphylatoxin C3a receptor (SEQ ID NO:2).

Figure 2A-2B shows design of the targeting construct used to disrupt anaphylatoxin C3a receptor genes. Figure 2B shows the sequences identified as SEQ ID NO:3 and SEQ ID NO:4, which were used as the targeting arms (homologous sequences) in the anaphylatoxin C3a receptor targeting construct.

Figure 3 shows a graph relating to thymus weight and body weight ratio of the wild-type mice and heterozygous mice and homozygous mice having a disruptions in anaphylatoxin C3a receptor genes.

Figure 4 shows the polynucleotide sequence for an 5-HT5A gene (SEQ ID NO:5). Figure 1 also shows the amino acid sequence for the an 5-HT5A gene (SEQ ID NO:6).

Figure 5A-5B shows design of the targeting construct used to disrupt 5-HT5A genes. Figure 5B shows the sequences identified as SEQ ID NO:7 and SEQ ID NO:8, which were used as the targeting arms (homologous sequences) in the 5-HT5A targeting construct.

Figure 6 shows the polynucleotide sequence for an orphan chordin (SEQ ID NO:9). Figure 1 also shows the amino acid sequence for a chordin polypeptide (SEQ ID NO:10).

Figure 7A-7B shows design of the targeting construct used to disrupt chordin genes. Figure 6B shows the sequences identified as SEQ ID NO:11 and SEQ ID NO:12, which were used as the targeting arms (homologous sequences) in the chordin targeting construct.

Figure 8 shows the polynucleotide sequence for an orphan RORγ (SEQ ID NO:13). Figure 8 also shows the amino acid sequence for the RORγ (SEQ ID NO:14).

Figure 9A-9B shows design of the targeting construct used to disrupt RORγ genes. Figure 9B shows the sequences identified as SEQ ID NO:15 and SEQ ID NO:16, which were used as the targeting arms (homologous sequences) in the RORγ targeting construct.

Figure 10 shows a graph relating to spleen weight and body weight ratio of the wild-type mice and heterozygous mice and homozygous mice having a disruption in RORγ genes.

Figure 11 shows a graph relating to liver weight and body weight ratio of the wild-type mice and heterozygous mice and homozygous mice having a disruption.

Figure 12 shows a graph relating to kidney weight and body weight ratio of the wild-type mice and heterozygous mice and homozygous mice having a disruption in RORγ genes.

Figure 13 shows a graph relating to thymus weight and body weight ratio of the wild-type mice and heterozygous mice and homozygous mice having a disruption in RORγ genes.

Figure 14 shows the polynucleotide sequence for a BMP gene (SEQ ID NO:17). Figure 14 also shows the amino acid sequence for a BMP polypeptide (SEQ ID NO:18).

Figure 15A-15B shows design of the targeting construct used to disrupt BMP genes. Figure 15B shows the sequences identified as SEQ ID NO:19 and SEQ ID NO:20, which were used as the targeting arms (homologous sequences) in the BMP targeting construct.

Figure 16 shows the polynucleotide sequence for an orphan airway trypsin-like protease (SEQ ID NO:21). Figure 16 also shows the amino acid sequence for the airway trypsin-like protease (SEQ ID NO:22).

Figure 17A-17B shows design of the targeting construct used to disrupt airway trypsin-like protease genes. Figure 17B shows the sequences identified as SEQ ID NO:23 and SEQ ID NO:24, which were used as the targeting arms (homologous sequences) in the airway trypsin-like protease targeting construct.

Figure 18 shows the polynucleotide sequence for an aquaporin-8 (SEQ ID NO:25). Figure 1 also shows the amino acid sequence for the aquaporin-8 (SEQ ID NO:26).

Figure 19A-19B shows design of the targeting construct used to disrupt aquaporin-8 genes. Figure 19B shows the sequences identified as SEQ ID NO:27 and SEQ ID NO:28, which were used as the targeting arms (homologous sequences) in the aquaporin-8 targeting construct.

Detailed Description of the Invention

The invention is based, in part, on the evaluation of the expression and role of genes and gene expression products, primarily those associated with a target gene. Among others, the invention permits the definition of disease pathways and the identification of diagnostically and therapeutically useful targets. For example, genes which are mutated or down-regulated under disease conditions may be involved in causing or exacerbating the disease condition. Treatments directed at up-regulating the activity of such genes or treatments which involve alternate pathways, may ameliorate the disease condition.

Generation of Targeting Construct

The targeting construct of the present invention may be produced using standard methods known in the art. (*See, e.g.*, Sambrook, *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; E.N. Glover (eds.), 1985, *DNA Cloning: A Practical Approach*, Volumes I and II; M.J. Gait (ed.), 1984,

Oligonucleotide Synthesis; B.D. Hames & S.J. Higgins (eds.), 1985, Nucleic Acid Hybridization; B.D. Hames & S.J. Higgins (eds.), 1984, *Transcription and Translation*; R.I. Freshney (ed.), 1986, Animal Cell Culture; Immobilized Cells and Enzymes, IRL Press, 1986; B. Perbal, 1984, A Practical Guide To Molecular Cloning; F.M. Ausubel *et al.*, 1994, *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.). For example, the targeting construct may be prepared in accordance with conventional ways, where sequences may be synthesized, isolated from natural sources, manipulated, cloned, ligated, subjected to *in vitro* mutagenesis, primer repair, or the like. At various stages, the joined sequences may be cloned, and analyzed by restriction analysis, sequencing, or the like.

The targeting DNA can be constructed using techniques well known in the art. For example, the targeting DNA may be produced by chemical synthesis of oligonucleotides, nick-translation of a double-stranded DNA template, polymerase chain-reaction amplification of a sequence (or ligase chain reaction amplification), purification of prokaryotic or target cloning vectors harboring a sequence of interest (*e.g.*, a cloned cDNA or genomic DNA, synthetic DNA or from any of the aforementioned combination) such as plasmids, phagemids, YACs, cosmids, bacteriophage DNA, other viral DNA or replication intermediates, or purified restriction fragments thereof, as well as other sources of single and double-stranded polynucleotides having a desired nucleotide sequence. Moreover, the length of homology may be selected using known methods in the art. For example, selection may be based on the sequence composition and complexity of the predetermined endogenous target DNA sequence(s).

The targeting construct of the present invention typically comprises a first sequence homologous to a portion or region of the target gene and a second sequence homologous to a second portion or region of the target gene. The targeting construct further comprises a positive selection marker, which is preferably positioned in between the first and the second DNA sequence that are homologous to a portion or region of the target DNA sequence. The positive selection marker may be operatively linked to a promoter and a polyadenylation signal.

Other regulatory sequences known in the art may be incorporated into the targeting construct to disrupt or control expression of a particular gene in a specific cell type. In addition, the targeting construct may also include a sequence coding for a screening marker, for example, green fluorescent protein (GFP), or another modified fluorescent protein.

Although the size of the homologous sequence is not critical and can range from as few as 50 base pairs to as many as 100 kb, preferably each fragment is greater than about 1 kb in length, more preferably between about 1 and about 10 kb, and even more preferably between about 1 and about 5 kb. One of skill in the art will recognize that although larger fragments may increase the number of homologous recombination events in ES cells, larger fragments will also be more difficult to clone.

In a preferred embodiment of the present invention, the targeting construct is prepared directly from a plasmid genomic library using the methods described in pending U.S. Patent Applica-

tion No.: 08/971,310, filed November 17, 1997, the disclosure of which is incorporated herein in its entirety. Generally, a sequence of interest is identified and isolated from a plasmid library in a single step using, for example, long-range PCR. Following isolation of this sequence, a second polynucleotide that will disrupt the target sequence can be readily inserted between two regions encoding the sequence of interest. In accordance with this aspect, the construct is generated in two steps by (1) amplifying (for example, using long-range PCR) sequences homologous to the target sequence, and (2) inserting another polynucleotide (for example a selectable marker) into the PCR product so that it is flanked by the homologous sequences. Typically, the vector is a plasmid from a plasmid genomic library. The completed construct is also typically a circular plasmid.

In another embodiment, the targeting construct is designed in accordance with the regulated positive selection method described in U.S. Application No. 60/232,957, filed September 15, 2000, the disclosure of which is incorporated herein in its entirety. The targeting construct is designed to include a PGK-*neo* fusion gene having two *lacO* sites, positioned in the PGK promoter and an NLS-*lacI* gene comprising a lac repressor fused to sequences encoding the NLS from the SV40 T antigen.

In another embodiment, the targeting construct may contain more than one selectable maker gene, including a negative selectable marker, such as the herpes simplex virus tk (HSV-tk) gene. The negative selectable marker may be operatively linked to a promoter and a polyadenylation signal. (*See, e.g.*, U.S. Patent No. 5,464,764; U.S. Patent No. 5,487,992; U.S. Patent No. 5,627,059; and U.S. Patent No. 5,631,153).

Generation of Cells and Confirmation of Homologous Recombination Events

Once an appropriate targeting construct has been prepared, the targeting construct may be introduced into an appropriate host cell using any method known in the art. Various techniques may be employed in the present invention, including, for example, pronuclear microinjection; retrovirus mediated gene transfer into germ lines; gene targeting in embryonic stem cells; electroporation of embryos; sperm-mediated gene transfer; and calcium phosphate/DNA co-precipitates, microinjection of DNA into the nucleus, bacterial protoplast fusion with intact cells, transfection, polycations, *e.g.*, polybrene, polyornithine, *etc.*, or the like (*See, e.g.*, U.S. Pat. No. 4,873,191; Van der Putten, *et al.*, 1985, *Proc. Natl. Acad. Sci., USA* 82:6148-6152; Thompson, *et al.*, 1989, *Cell* 56:313-321; Lo, 1983, *Mol Cell. Biol.* 3:1803-1814; Lavitrano, *et al.*, 1989, *Cell*, 57:717-723). Various techniques for transforming mammalian cells are known in the art. (*See, e.g.*, Gordon, 1989, *Intl. Rev. Cytol.*, 115:171-229; Keown *et al.*, 1989, *Methods in Enzymology*; Keown *et al.*, 1990, *Methods and Enzymology*, Vol. 185, pp. 527-537; Mansour *et al.*, 1988, *Nature*, 336:348-352).

In a preferred aspect of the present invention, the targeting construct is introduced into host cells by electroporation. In this process, electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the construct. The pores created during

electroporation permit the uptake of macromolecules such as DNA. (*See, e.g.*, Potter, H., *et al.*, 1984, *Proc. Nat'l. Acad. Sci. U.S.A.* 81:7161-7165).

Any cell type capable of homologous recombination may be used in the practice of the present invention. Examples of such target cells include cells derived from vertebrates including mammals such as humans, bovine species, ovine species, murine species, simian species, and ether eucaryotic organisms such as filamentous fungi, and higher multicellular organisms such as plants.

Preferred cell types include embryonic stem (ES) cells, which are typically obtained from pre-implantation embryos cultured *in vitro*. (*See, e.g.*, Evans, M. J., *et al.*, 1981, *Nature* 292:154-156; Bradley, M. O., *et al.*, 1984, *Nature* 309:255-258; Gossler *et al.*, 1986, *Proc. Natl. Acad. Sci. USA* 83:9065-9069; and Robertson, *et al.*, 1986, *Nature* 322:445-448). The ES cells are cultured and prepared for introduction of the targeting construct using methods well known to the skilled artisan. (*See, e.g.*, Robertson, E. J. ed. "Teratocarcinomas and Embryonic Stem Cells, a Practical Approach", IRL Press, Washington D.C., 1987; Bradley *et al.*, 1986, *Current Topics in Devel. Biol.* 20:357-371; by Hogan *et al.* in "Manipulating the Mouse Embryo": A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y., 1986; Thomas *et al.*, 1987, *Cell* 51:503; Koller *et al.*, 1991, *Proc. Natl. Acad. Sci. USA*, 88:10730; Dorin *et al.*, 1992, *Transgenic Res.* 1:101; and Veis *et al.*, 1993, *Cell* 75:229). The ES cells that will be inserted with the targeting construct are derived from an embryo or blastocyst of the same species as the developing embryo into which they are to be introduced. ES cells are typically selected for their ability to integrate into the inner cell mass and contribute to the germ line of an individual when introduced into the mammal in an embryo at the blastocyst stage of development. Thus, any ES cell line having this capability is suitable for use in the practice of the present invention.

The present invention may also be used to knockout genes in other cell types, such as stem cells. By way of example, stem cells may be myeloid, lymphoid, or neural progenitor and precursor cells. These cells comprising a disruption or knockout of a gene may be particularly useful in the study of target gene function in individual developmental pathways. Stem cells may be derived from any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the like.

After the targeting construct has been introduced into cells, the cells where successful gene targeting has occurred are identified. Insertion of the targeting construct into the targeted gene is typically detected by identifying cells for expression of the marker gene. In a preferred embodiment, the cells transformed with the targeting construct of the present invention are subjected to treatment with an appropriate agent that selects against cells not expressing the selectable marker. Only those cells expressing the selectable marker gene survive and/or grow under certain conditions. For example, cells that express the introduced neomycin resistance gene are resistant to the compound G418, while cells that do not express the neo gene marker are killed by G418. If the targeting

construct also comprises a screening marker such as GFP, homologous recombination can be identified through screening cell colonies under a fluorescent light. Cells that have undergone homologous recombination will have deleted the GFP gene and will not fluoresce.

If a regulated positive selection method is used in identifying homologous recombination events, the targeting construct is designed so that the expression of the selectable marker gene is regulated in a manner such that expression is inhibited following random integration but is permitted (derepressed) following homologous recombination. More particularly, the transfected cells are screened for expression of the *neo* gene, which requires that (1) the cell was successfully electroporated, and (2) *lac* repressor inhibition of *neo* transcription was relieved by homologous recombination. This method allows for the identification of transfected cells and homologous recombinants to occur in one step with the addition of a single drug.

Alternatively, a positive-negative selection technique may be used to select homologous recombinants. This technique involves a process in which a first drug is added to the cell population, for example, a neomycin-like drug to select for growth of transfected cells, *i.e.* positive selection. A second drug, such as FIAU is subsequently added to kill cells that express the negative selection marker, *i.e.* negative selection. Cells that contain and express the negative selection marker are killed by a selecting agent, whereas cells that do not contain and express the negative selection marker survive. For example, cells with non-homologous insertion of the construct express HSV thymidine kinase and therefore are sensitive to the herpes drugs such as gancyclovir (GANC) or FIAU (1-(2-deoxy 2-fluoro-B-D-arabinofluranosyl)-5-iodouracil). (*See, e.g.*, Mansour *et al.*, *Nature* 336:348-352: (1988); Capecchi, *Science* 244:1288-1292, (1989); Capecchi, *Trends in Genet.* 5:70-76 (1989)).

Successful recombination may be identified by analyzing the DNA of the selected cells to confirm homologous recombination. Various techniques known in the art, such as PCR and/or Southern analysis may be used to confirm homologous recombination events.

Homologous recombination may also be used to disrupt genes in stem cells, and other cell types, which are not totipotent embryonic stem cells. By way of example, stem cells may be myeloid, lymphoid, or neural progenitor and precursor cells. Such transgenic cells may be particularly useful in the study of gene function in individual developmental pathways. Stem cells may be derived from any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the like.

In cells which are not totipotent it may be desirable to knock out both copies of the target using methods which are known in the art. For example, cells comprising homologous recombination at a target locus which have been selected for expression of a positive selection marker (e.g., Neo^r) and screened for non-random integration, can be further selected for multiple copies of the selectable marker gene by exposure to elevated levels of the selective agent (e.g., G418). The cells are then analyzed for homozygosity at the target locus. Alternatively, a second construct can be generated

with a different positive selection marker inserted between the two homologous sequences. The two constructs can be introduced into the cell either sequentially or simultaneously, followed by appropriate selection for each of the positive marker genes. The final cell is screened for homologous recombination of both alleles of the target.

Production of Transgenic Animals

Selected cells are then injected into a blastocyst (or other stage of development suitable for the purposes of creating a viable animal, such as, for example, a morula) of an animal (*e.g.*, a mouse) to form chimeras (*see e.g.*, Bradley, A. *in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed., IRL, Oxford, pp. 113-152 (1987)). Alternatively, selected ES cells can be allowed to aggregate with dissociated mouse embryo cells to form the aggregation chimera. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Chimeric progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA. In one embodiment, chimeric progeny mice are used to generate a mouse with a heterozygous disruption in the gene. Heterozygous transgenic mice can then be mated. It is well know in the art that typically ¼ of the offspring of such matings will have a homozygous disruption in the target gene.

The heterozygous and homozygous transgenic mice can then be compared to normal, wild type mice to determine whether disruption of the target gene causes phenotypic changes, especially pathological changes. For example, heterozygous and homozygous mice may be evaluated for phenotypic changes by physical examination, necropsy, histology, clinical chemistry, complete blood count, body weight, organ weights, and cytological evaluation of bone marrow.

In one embodiment, the phenotype (or phenotypic change) associated with a disruption in the target gene is placed into or stored in a database. Preferably, the database includes: (i) genotypic data (e.g., identification of the disrupted gene) and (ii) phenotypic data (e.g., phenotype(s) resulting from the gene disruption) associated with the genotypic data. The database is preferably electronic. In addition, the database is preferably combined with a search tool so that the database is searchable.

Conditional Transgenic Animals

The present invention further contemplates conditional transgenic or knockout animals, such as those produced using recombination methods. Bacteriophage P1 Cre recombinase and flp recombinase from yeast plasmids are two non-limiting examples of site-specific DNA recombinase enzymes which cleave DNA at specific target sites (lox P sites for cre recombinase and frt sites for flp recombinase) and catalyze a ligation of this DNA to a second cleaved site. A large number of suitable alternative site-specific recombinases have been described, and their genes can be used in accordance with the method of the present invention. Such recombinases include the Int recombinase of bacteriophage λ (with or without Xis) (Weisberg, R. et. al., in *Lambda II*, (Hendrix, R., *et al.*, Eds.), Cold

Spring Harbor Press, Cold Spring Harbor, NY, pp. 211-50 (1983), herein incorporated by reference); TpnI and the β-lactamase transposons (Mercier, *et al.*, *J. Bacteriol.*, 172:3745-57 (1990)); the Tn3 resolvase (Flanagan & Fennewald *J. Molec. Biol.*, 206:295-304 (1989); Stark, *et al.*, *Cell*, 58:779-90 (1989)); the yeast recombinases (Matsuzaki, *et al.*, *J. Bacteriol.*, 172:610-18 (1990)); the *B. subtilis* SpoIVC recombinase (Sato, *et al.*, *J. Bacteriol.* 172:1092-98 (1990)); the Flp recombinase (Schwartz & Sadowski, *J. Molec.Biol.*, 205:647-658 (1989); Parsons, *et al.*, *J. Biol. Chem.*, 265:4527-33 (1990); Golic & Lindquist, *Cell*, 59:499-509 (1989); Amin, *et al.*, *J. Molec. Biol.*, 214:55-72 (1990)); the Hin recombinase (Glasgow, *et al.*, *J. Biol. Chem.*, 264:10072-82 (1989)); immunoglobulin recombinases (Malynn, *et al.*, *Cell*, 54:453-460 (1988)); and the Cin recombinase (Haffter & Bickle, *EMBO J.*, 7:3991-3996 (1988); Hubner, *et al.*, *J. Molec. Biol.*, 205:493-500 (1989)), all herein incorporated by reference. Such systems are discussed by Echols (*J. Biol. Chem.* 265:14697-14700 (1990)); de Villartay (*Nature*, 335:170-74 (1988)); Craig, (*Ann. Rev. Genet.*, 22:77-105 (1988)); Poyart-Salmeron, *et al.*, (*EMBO J.* 8:2425-33 (1989)); Hunger-Bertling, *et al.* (*Mol Cell. Biochem.*, 92:107-16 (1990)); and Cregg & Madden (*Mol. Gen. Genet.*, 219:320-23 (1989)), all herein incorporated by reference.

Cre has been purified to homogeneity, and its reaction with the loxP site has been extensively characterized (Abremski & Hess *J. Mol. Biol.* 259:1509-14 (1984), herein incorporated by reference). Cre protein has a molecular weight of 35,000 and can be obtained commercially from New England Nuclear/DuPont. The cre gene (which encodes the Cre protein) has been cloned and expressed (Abremski, *et al. Cell* 32:1301-11 (1983), herein incorporated by reference). The Cre protein mediates recombination between two loxP sequences (Sternberg, *et al. Cold Spring Harbor Symp. Quant. Biol.* 45:297-309 (1981)), which may be present on the same or different DNA molecule. Because the internal spacer sequence of the loxP site is asymmetrical, two loxP sites can exhibit directionality relative to one another (Hoess & Abremski *Proc. Natl. Acad. Sci. U.S.A.* 81:1026-29 (1984)). Thus, when two sites on the same DNA molecule are in a directly repeated orientation, Cre will excise the DNA between the sites (Abremski, *et al. Cell* 32:1301-11 (1983)). However, if the sites are inverted with respect to each other, the DNA between them is not excised after recombination but is simply inverted. Thus, a circular DNA molecule having two loxP sites in direct orientation will recombine to produce two smaller circles, whereas circular molecules having two loxP sites in an inverted orientation simply invert the DNA sequences flanked by the loxP sites. In addition, recombinase action can result in reciprocal exchange of regions distal to the target site when targets are present on separate DNA molecules.

Recombinases have important application for characterizing gene function in knockout models. When the constructs described herein are used to disrupt target genes, a fusion transcript can be produced when insertion of the positive selection marker occurs downstream (3') of the translation initiation site of the target gene. The fusion transcript could result in some level of protein expression with unknown consequence. It has been suggested that insertion of a positive selection marker gene

can affect the expression of nearby genes. These effects may make it difficult to determine gene function after a knockout event since one could not discern whether a given phenotype is associated with the inactivation of a gene, or the transcription of nearby genes. Both potential problems are solved by exploiting recombinase activity. When the positive selection marker is flanked by recombinase sites in the same orientation, the addition of the corresponding recombinase will result in the removal of the positive selection marker. In this way, effects caused by the positive selection marker or expression of fusion transcripts are avoided.

In one embodiment, purified recombinase enzyme is provided to the cell by direct microinjection. In another embodiment, recombinase is expressed from a co-transfected construct or vector in which the recombinase gene is operably linked to a functional promoter. An additional aspect of this embodiment is the use of tissue-specific or inducible recombinase constructs which allow the choice of when and where recombination occurs. One method for practicing the inducible forms of recombinase-mediated recombination involves the use of vectors that use inducible or tissue-specific promoters or other gene regulatory elements to express the desired recombinase activity. The inducible expression elements are preferably operatively positioned to allow the inducible control or activation of expression of the desired recombinase activity. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No, *et al. Proc. Natl. Acad. Sci. USA*, 93:3346-51 (1996); Furth, *et al. Proc. Natl. Acad. Sci. USA*, 91:9302-6 (1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, promoters. Vectors incorporating such promoters would only express recombinase activity in cells that express the necessary transcription factors.

Models for Disease

The cell- and animal-based systems described herein can be utilized as models for diseases. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate disease animal models. In addition, cells from humans may be used. These systems may be used in a variety of applications. Such assays may be utilized as part of screening strategies designed to identify agents, such as compounds which are capable of ameliorating disease symptoms. Thus, the animal- and cell-based models may be used to identify drugs, pharmaceuticals, therapies and interventions which may be effective in treating disease.

Cell-based systems may be used to identify compounds which may act to ameliorate disease symptoms. For example, such cell systems may be exposed to a compound suspected of exhibiting an ability to ameliorate disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed cells. After exposure, the cells are

examined to determine whether one or more of the disease cellular phenotypes has been altered to resemble a more normal or more wild type, non-disease phenotype.

In addition, animal-based disease systems, such as those described herein, may be used to identify compounds capable of ameliorating disease symptoms. Such animal models may be used as test substrates for the identification of drugs, pharmaceuticals, therapies, and interventions which may be effective in treating a disease or other phenotypic characteristic of the animal. For example, animal models may be exposed to a compound or agent suspected of exhibiting an ability to ameliorate disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed animals. The response of the animals to the exposure may be monitored by assessing the reversal of disorders associated with the disease. Exposure may involve treating mother animals during gestation of the model animals described herein, thereby exposing embryos or fetuses to the compound or agent which may prevent or ameliorate the disease or phenotype. Neonatal, juvenile, and adult animals can also be exposed.

More particularly, using the animal models of the invention, specifically, transgenic mice, methods of identifying agents, including compounds are provided, preferably, on the basis of the ability to affect at least one phenotype associated with a disruption in a target gene. In one embodiment, the present invention provides a method of identifying agents having an effect on target gene or alternatively, target protein expression or function. The method includes measuring a physiological response of the animal, for example, to the agent, and comparing the physiological response of such animal to a control animal, wherein the physiological response of the animal comprising a disruption in a target gene as compared to the control animal indicates the specificity of the agent. A "physiological response" is any biological or physical parameter of an animal which can be measured. Molecular assays (e.g., gene transcription, protein production and degradation rates), physical parameters (e.g., exercise physiology tests, measurement of various parameters of respiration, measurement of heart rate or blood pressure, measurement of bleeding time, aPTT.T, or TT), and cellular assays (e.g.,. immunohistochemical assays of cell surface markers, or the ability of cells to aggregate or proliferate) can be used to assess a physiological response. The transgenic animals and cells of the present invention may by utilized as models for diseases, disorders, or conditions associated with phenotypes relating to a disruption in a target gene.

The present invention provides a unique animal model for testing and developing new treatments relating to the behavioral phenotypes. Analysis of the behavioral phenotype allows for the development of an animal model useful for testing, for instance, the efficacy of proposed genetic and pharmacological therapies for human genetic diseases, such as neurological, neuropsychological, or psychotic illnesses.

A statistical analysis of the various behaviors measured can be carried out using any conventional statistical program routinely used by those skilled in the art (such as, for example,

"Analysis of Variance" or ANOVA). A "p" value of about 0.05 or less is generally considered to be statistically significant, although slightly higher p values may still be indicative of statistically significant differences. To statistically analyze abnormal behavior, a comparison is made between the behavior of a transgenic animal (or a group thereof) to the behavior of a wild-type mouse (or a group thereof), typically under certain prescribed conditions. "Abnormal behavior" as used herein refers to behavior exhibited by an animal having a disruption in the target gene, e.g. transgenic animal, which differs from an animal without a disruption in the target gene, *e.g.* wild-type mouse. Abnormal behavior consists of any number of standard behaviors that can be objectively measured (or observed) and compared. In the case of comparison, it is preferred that the change be statistically significant to confirm that there is indeed a meaningful behavioral difference between the knockout animal and the wild-type control animal. Examples of behaviors which may be measured or observed include, but are not limited to, ataxia, rapid limb movement, eye movement, breathing, motor activity, cognition, emotional behaviors, social behaviors, hyperactivity, hypersensitivity, anxiety, impaired learning, abnormal reward behavior, and abnormal social interaction, such as aggression.

A series of tests may be used to measure the behavioral phenotype of the animal models of the present invention, including neurological and neuropsychological tests to identify abnormal behavior. These tests may be used to measure abnormal behavior relating to, for example, learning and memory, eating, pain, aggression, sexual reproduction, anxiety, depression, schizophrenia, and drug abuse. (*See, e.g.*, Crawley and Paylor, *Hormones and Behavior* 31:197-211 (1997)).

The social interaction test involves exposing a mouse to other animals in a variety of settings. The social behaviors of the animals (e.g., touching, climbing, sniffing, and mating) are subsequently evaluated. Differences in behaviors can then be statistically analyzed and compared (*See, e.g.*, S. E. File, *et al.*, *Pharmacol. Bioch. Behav.* 22:941-944 (1985); R. R. Holson, *Phys. Behav.* 37:239-247 (1986)). Examplary behavioral tests include the following.

The mouse startle response test typically involves exposing the animal to a sensory (typically auditory) stimulus and measuring the startle response of the animal (*see, e.g.*, M. A. Geyer, *et al.*, *Brain Res. Bull.* 25:485-498 (1990); Paylor and Crawley, *Psychopharmacology* 132:169-180 (1997)). A pre-pulse inhibition test can also be used, in which the percent inhibition (from a normal startle response) is measured by "cueing" the animal first with a brief low-intensity pre-pulse prior to the startle pulse.

The electric shock test generally involves exposure to an electrified surface and measurement of subsequent behaviors such as, for example, motor activity, learning, social behaviors. The behaviors are measured and statistically analyzed using standard statistical tests. (*See, e.g.*, G. J. Kant, *et al.*, *Pharm. Bioch. Behav.* 20:793-797 (1984); N. J. Leidenheimer, *et al.*, *Pharmacol. Bioch. Behav.* 30:351-355 (1988)).

The tail-pinch or immobilization test involves applying pressure to the tail of the animal and/or restraining the animal's movements. Motor activity, social behavior, and cognitive behavior are examples of the areas that are measured. (*See, e.g.*, M. Bertolucci D'Angic, *et al.*, *Neurochem.* 55:1208-1214 (1990)).

The novelty test generally comprises exposure to a novel environment and/or novel objects. The animal's motor behavior in the novel environment and/or around the novel object are measured and statistically analyzed. (*See, e.g.*, D. K. Reinstein, *et al.*, *Pharm. Bioch. Behav.* 17:193-202 (1982); B. Poucet, *Behav. Neurosci.* 103:1009-10016 (1989); R. R. Holson, *et al.*, *Phys. Behav.* 37:231-238 (1986)). This test may be used to detect visual processing deficiencies or defects.

The learned helplessness test involves exposure to stresses, for example, noxious stimuli, which cannot be affected by the animal's behavior. The animal's behavior can be statistically analyzed using various standard statistical tests. (*See, e.g.*, A. Leshner, *et al.*, *Behav. Neural Biol.* 26:497-501 (1979)).

Alternatively, a tail suspension test may be used, in which the "immobile" time of the mouse is measured when suspended "upside-down" by its tail. This is a measure of whether the animal struggles, an indicator of depression. In humans, depression is believed to result from feelings of a lack of control over one's life or situation. It is believed that a depressive state can be elicited in animals by repeatedly subjecting them to aversive situations over which they have no control. A condition of "learned helplessness" is eventually reached, in which the animal will stop trying to change its circumstances and simply accept its fate. Animals that stop struggling sooner are believed to be more prone to depression. Studies have shown that the administration of certain antidepressant drugs prior to testing increases the amount of time that animals struggle before giving up.

The Morris water-maze test comprises learning spatial orientations in water and subsequently measuring the animal's behaviors, such as, for example, by counting the number of incorrect choices. The behaviors measured are statistically analyzed using standard statistical tests. (*See, e.g.*, E. M. Spruijt, *et al.*, *Brain Res.* 527:192-197 (1990)).

Alternatively, a Y-shaped maze may be used (see, e.g., McFarland, D.J., *Pharmacology, Biochemistry and Behavior* 32:723-726 (1989); Dellu, F., et al., *Neurobiology of Learning and Memory* 73:31-48 (2000)). The Y-maze is generally believed to be a test of cognitive ability. The dimensions of each arm of the Y-maze can be, for example, approximately 40 cm x 8 cm x 20 cm, although other dimensions may be used. Each arm can also have, for example, sixteen equally spaced photobeams to automatically detect movement within the arms. At least two different tests can be performed using such a Y-maze. In a continuous Y-maze paradigm, mice are allowed to explore all three arms of a Y-maze for, e.g., approximately 10 minutes. The animals are continuously tracked using photobeam detection grids, and the data can be used to measure spontaneous alteration and positive bias behavior. Spontaneous alteration refers to the natural tendency of a "normal" animal to

visit the least familiar arm of a maze. An alternation is scored when the animal makes two consecutive turns in the same direction, thus representing a sequence of visits to the least recently entered arm of the maze. Position bias determines egocentrically defined responses by measuring the animal's tendency to favor turning in one direction over another. Therefore, the test can detect differences in an animal's ability to navigate on the basis of allocentric or egocentric mechanisms. The two-trial Y-maze memory test measures response to novelty and spatial memory based on a free-choice exploration paradigm. During the first trial (acquisition), the animals are allowed to freely visit two arms of the Y-maze for, e.g., approximately 15 minutes. The third arm is blocked off during this trial. The second trial (retrieval) is performed after an intertrial interval of, e.g., approximately 2 hours. During the retrieval trial, the blocked arm is opened and the animal is allowed access to all three arms for, e.g., approximately 5 minutes. Data are collected during the retrieval trial and analyzed for the number and duration of visits to each arm. Because the three arms of the maze are virtually identical, discrimination between novelty and familiarity is dependent on "environmental" spatial cues around the room relative to the position of each arm. Changes in arm entry and duration of time spent in the novel arm in a transgenic animal model may be indicative of a role of that gene in mediating novelty and recognition processes.

The passive avoidance or shuttle box test generally involves exposure to two or more environments, one of which is noxious, providing a choice to be learned by the animal. Behavioral measures include, for example, response latency, number of correct responses, and consistency of response. (*See, e.g.*, R. Ader, *et al.*, *Psychon. Sci.* 26:125-128 (1972); R. R. Holson, *Phys. Behav.* 37:221-230 (1986)). Alternatively, a zero-maze can be used. In a zero-maze, the animals can, for example, be placed in a closed quadrant of an elevated annular platform having, e.g., 2 open and 2 closed quadrants, and are allowed to explore for approximately 5 minutes. This paradigm exploits an approach-avoidance conflict between normal exploratory activity and an aversion to open spaces in rodents. This test measures anxiety levels and can be used to evaluate the effectiveness of anti-anxiolytic drugs. The time spent in open quadrants versus closed quadrants may be recorded automatically, with, for example, the placement of photobeams at each transition site.

The food avoidance test involves exposure to novel food and objectively measuring, for example, food intake and intake latency. The behaviors measured are statistically analyzed using standard statistical tests. (*See, e.g.*, B. A. Campbell, et al., *J. Comp. Physiol. Psychol.* 67:15-22 (1969)).

The elevated plus-maze test comprises exposure to a maze, without sides, on a platform, the animal's behavior is objectively measured by counting the number of maze entries and maze learning. The behavior is statistically analyzed using standard statistical tests. (*See, e.g.*, H. A. Baldwin, *et al.*, *Brain Res. Bull*, 20:603-606 (1988)).

The stimulant-induced hyperactivity test involves injection of stimulant drugs (*e.g.*, amphetamines, cocaine, PCP, and the like), and objectively measuring, for example, motor activity, social interactions, cognitive behavior. The animal's behaviors are statistically analyzed using standard statistical tests. (*See, e.g.*, P. B. S. Clarke, *et al.*, *Psychopharmacology* 96:511-520 (1988); P. Kuczenski, *et al.*, *J. Neuroscience* 11:2703-2712 (1991)).

The self-stimulation test generally comprises providing the mouse with the opportunity to regulate electrical and/or chemical stimuli to its own brain. Behavior is measured by frequency and pattern of self-stimulation. Such behaviors are statistically analyzed using standard statistical tests. (*See, e.g.*, S. Nassif, *et al.*, *Brain Res.*, 332:247-257 (1985); W. L. Isaac, *et al.*, *Behav. Neurosci.* 103:345-355 (1989)).

The reward test involves shaping a variety of behaviors, e.g., motor, cognitive, and social, measuring, for example, rapidity and reliability of behavioral change, and statistically analyzing the behaviors measured. (*See, e.g.*, L. E. Jarrard, *et al.*, *Exp. Brain Res.* 61:519-530 (1986)).

The DRL (differential reinforcement to low rates of responding) performance test involves exposure to intermittent reward paradigms and measuring the number of proper responses, e.g., lever pressing. Such behavior is statistically analyzed using standard statistical tests. (*See, e.g.*, J. D. Sinden, *et al.*, *Behav. Neurosci.* 100:320-329 (1986); V. Nalwa, *et al.*, *Behav Brain Res.* 17:73-76 (1985); and A. J. Nonneman, *et al.*, *J. Comp. Physiol. Psych.* 95:588-602 (1981)).

The spatial learning test involves exposure to a complex novel environment, measuring the rapidity and extent of spatial learning, and statistically analyzing the behaviors measured. (*See, e.g.*, N. Pitsikas, *et al.*, *Pharm. Bioch. Behav.* 38:931-934 (1991); B. Poucet, *et al.*, *Brain Res.* 37:269-280 (1990); D. Christie, *et al.*, *Brain Res.* 37:263-268 (1990); and F. Van Haaren, *et al.*, *Behav. Neurosci.* 102:481-488 (1988)). Alternatively, an open-field (of) test may be used, in which the greater distance traveled for a given amount of time is a measure of the activity level and anxiety of the animal. When the open field is a novel environment, it is believed that an approach-avoidance situation is created, in which the animal is "torn" between the drive to explore and the drive to protect itself. Because the chamber is lighted and has no places to hide other than the corners, it is expected that a "normal" mouse will spend more time in the corners and around the periphery than it will in the center where there is no place to hide. "Normal" mice will, however, venture into the central regions as they explore more and more of the chamber. It can then be extrapolated that especially anxious mice will spend most of their time in the corners, with relatively little or no exploration of the central region, whereas bold (i.e., less anxious) mice will travel a greater distance, showing little preference for the periphery versus the central region.

The visual, somatosensory and auditory neglect tests generally comprise exposure to a sensory stimulus, objectively measuring, for example, orientating responses, and statistically analyzing the behaviors measured. (See, e.g., J. M. Vargo, et al., Exp. Neurol. 102:199-209 (1988)).

The consummatory behavior test generally comprises feeding and drinking, and objectively measuring quantity of consumption. The behavior measured is statistically analyzed using standard statistical tests. (See, e.g., P. J. Fletcher, et al., Psychopharmacol. 102:301-308 (1990); M. G. Corda, et al.,, Proc. Nat'l Acad. Sci. USA 80:2072-2076 (1983)).

A visual discrimination test can also be used to evaluate the visual processing of an animal. One or two similar objects are placed in an open field and the animal is allowed to explore for about 5-10 minutes. The time spent exploring each object (proximity to, i.e., movement within, e.g., about 3-5 cm of the object is considered exploration of an object) is recorded. The animal is then removed from the open field, and the objects are replaced by a similar object and a novel object. The animal is returned to the open field and the percent time spent exploring the novel object over the old object is measured (again, over about a 5-10 minute span). "Normal" animals will typically spend a higher percentage of time exploring the novel object rather than the old object. If a delay is imposed between sampling and testing, the memory task becomes more hippocampal-dependent. If no delay is imposed, the task is more based on simple visual discrimination. This test can also be used for olfactory discrimination, in which the objects (preferably, simple blocks) can be sprayed or otherwise treated to hold an odor. This test can also be used to determine if the animal can make gustatory discriminations; animals that return to the previously eaten food instead of novel food exhibit gustatory neophobia.

A hot plate analgesia test can be used to evaluate an animal's sensitivity to heat or painful stimuli. For example, a mouse can be placed on an approximately 55°C hot plate and the mouse's response latency (e.g., time to pick up and lick a hind paw) can be recorded. These responses are not reflexes, but rather "higher" responses requiring cortical involvement. This test may be used to evaluate a nociceptive disorder.

An accelerating rotarod test may be used to measure coordination and balance in mice. Animals can be, for example, placed on a rod that acts like a rotating treadmill (or rolling log). The rotarod can be made to rotate slowly at first and then progressively faster until it reaches a speed of, e.g., approximately 60 rpm. The mice must continually reposition themselves in order to avoid falling off. The animals are preferably tested in at least three trials, a minimum of 20 minutes apart. Those mice that are able to stay on the rod the longest are believed to have better coordination and balance.

A metrazol administration test can be used to screen animals for varying susceptibilities to seizures or similar events. For example, a 5mg/ml solution of metrazol can be infused through the tail vein of a mouse at a rate of, e.g., approximately 0.375 ml/min. The infusion will cause all mice to experience seizures, followed by death. Those mice that enter the seizure stage the soonest are believed to be more prone to seizures. Four distinct physiological stages can be recorded: soon after the start of infusion, the mice will exhibit a noticeable "twitch", followed by a series of seizures, ending in a final tensing of the body known as "tonic extension", which is followed by death.

Target Gene Products

The present invention further contemplates use of the target gene sequence to produce target gene products. Target gene products may include proteins that represent functionally equivalent gene products. Such an equivalent gene product may contain deletions, additions or substitutions of amino acid residues within the amino acid sequence encoded by the gene sequences described herein, but which result in a silent change, thus producing a functionally equivalent target gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Functionally equivalent", as utilized herein, refers to a protein capable of exhibiting a substantially similar in vivo activity as the endogenous gene products encoded by the target gene sequences. Alternatively, when utilized as part of an assay, "functionally equivalent" may refer to peptides capable of interacting with other cellular or extracellular molecules in a manner substantially similar to the way in which the corresponding portion of the endogenous gene product would.

Other protein products useful according to the methods of the invention are peptides derived from or based on the target gene produced by recombinant or synthetic means (derived peptides).

Target gene products may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing the gene polypeptides and peptides of the invention by expressing nucleic acid encoding gene sequences are described herein. Methods which are well known to those skilled in the art can be used to construct expression vectors containing gene protein coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination/genetic recombination (see, *e.g.*, Sambrook, et al., 1989, *supra*, and Ausubel, et al., 1989, supra). Alternatively, RNA capable of encoding gene protein sequences may be chemically synthesized using, for example, automated synthesizers (see, *e.g.* Oligonucleotide Synthesis: A Practical Approach, Gait, M. J. ed., IRL Press, Oxford (1984)).

A variety of host-expression vector systems may be utilized to express the gene coding sequences of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells which may, when transformed or transfected with the appropriate nucleotide coding sequences, exhibit the gene protein of the invention *in situ*. These include but are not limited to microorganisms such as bacteria (*e.g.*, E. coli, B. subtilis) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid

DNA expression vectors containing gene protein coding sequences; yeast (*e.g.* Saccharomyces, Pichia) transformed with recombinant yeast expression vectors containing the gene protein coding sequences; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the gene protein coding sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing gene protein coding sequences; or mammalian cell systems (*e.g.* COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionine promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5 K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the gene protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the E. coli expression vector pUR278 (Ruther *et al.*, *EMBO J.*, 2:1791-94 (1983)), in which the gene protein coding sequence may be ligated individually into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, *Nucleic Acids Res.*, 13:3101-09 (1985); Van Heeke *et al.*, *J. Biol. Chem.*, 264:5503-9 (1989)); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

In a preferred embodiment, full length cDNA sequences are appended with in-frame Bam HI sites at the amino terminus and Eco RI sites at the carboxyl terminus using standard PCR methodologies (Innis, et al. (eds) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego (1990)) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radioactive labeling and glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, *et al.*, *EMBO J.*, 4: 1075-80 (1985); Zabeau *et al.*, *EMBO J.*, 1: 1217-24 (1982)).

In an insect system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in Spodoptera frugiperda cells. The gene coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of gene coding sequence will result in inactivation of the polyhedrin gene and

production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect Spodoptera frugiperda cells in which the inserted gene is expressed (see, *e.g.*, Smith, *et al.*, *J. Virol.* 46: 584-93 (1983); U.S. Pat. No. 4,745,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the gene coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by in vitro or in vivo recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing gene protein in infected hosts. (*e.g.*, see Logan *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:3655-59 (1984)). Specific initiation signals may also be required for efficient translation of inserted gene coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the gene coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bitter, *et al.*, *Methods in Enzymol.*, 153:516-44 (1987)).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the gene protein may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences,

transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells which stably integrate the plasmid into their chromosomes and grow, to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the gene protein. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the gene protein.

In a preferred embodiment, control of timing and/or quantity of expression of the recombinant protein can be controlled using an inducible expression construct. Inducible constructs and systems for inducible expression of recombinant proteins will be well known to those skilled in the art. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No, *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:3346-51 (1996); Furth, *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:9302-6 (1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, particularly HIV, promoters. In one in embodiment, a Tet inducible gene expression system is utilized. (Gossen *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:5547-51 (1992); Gossen, *et al.*, *Science*, 268:1766-69 (1995)). Tet Expression Systems are based on two regulatory elements derived from the tetracycline-resistance operon of the *E. coli* Tn10 transposon—the tetracycline repressor protein (TetR) and the tetracycline operator sequence (tetO) to which TetR binds. Using such a system, expression of the recombinant protein is placed under the control of the tetO operator sequence and transfected or transformed into a host cell. In the presence of TetR, which is co-transfected into the host cell, expression of the recombinant protein is repressed due to binding of the TetR protein to the tetO regulatory element. High-level, regulated gene expression can then be induced in response to varying concentrations of tetracycline (Tc) or Tc derivatives such as doxycycline (Dox), which compete with tetO elements for binding to TetR. Constructs and materials for tet inducible gene expression are available commercially from CLONTECH Laboratories, Inc., Palo Alto, CA.

When used as a component in an assay system, the gene protein may be labeled, either directly or indirectly, to facilitate detection of a complex formed between the gene protein and a test substance. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as 125I; enzyme labeling systems that generate a detectable calorimetric signal or light when exposed to substrate; and fluorescent labels. Where recombinant DNA technology is used to produce the gene protein for such assay systems, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection.

Indirect labeling involves the use of a protein, such as a labeled antibody, which specifically binds to the gene product. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

Production of Antibodies

Described herein are methods for the production of antibodies capable of specifically recognizing one or more epitopes. Such antibodies may include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')2 fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. Such antibodies may be used, for example, in the detection of a target gene in a biological sample, or, alternatively, as a method for the inhibition of abnormal target gene activity. Thus, such antibodies may be utilized as part of disease treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels of target gene proteins, or for the presence of abnormal forms of such proteins.

For the production of antibodies, various host animals may be immunized by injection with the target gene, its expression product or a portion thereof. Such host animals may include but are not limited to rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (*bacille Calmette-Guerin*) and *Corynebacterium parvum*.

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as target gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with gene product supplemented with adjuvants as also described above.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to the hybridoma technique of Köhler and Milstein, *Nature*, 256:495-7 (1975); and U.S. Pat. No. 4,376,110), the human B-cell hybridoma technique (Kosbor, *et al.*, *Immunology Today*, 4:72 (1983); Cote, *et al.*, *Proc. Natl. Acad. Sci. USA*, 80:2026-30 (1983)), and the EBV-hybridoma technique (Cole, *et al.*, in Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., New York, pp. 77-96 (1985)). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The

hybridoma producing the mAb of this invention may be cultivated in vitro or in vivo. Production of high titers of mAbs in vivo makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison, *et al., Proc. Natl. Acad. Sci.*, 81:6851-6855 (1984); Takeda, *et al., Nature*, 314:452-54 (1985)) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778; Bird, *Science* 242:423-26 (1988); Huston, *et al., Proc. Natl. Acad. Sci. USA*, 85:5879-83 (1988); and Ward, *et al., Nature*, 334:544-46 (1989)) can be adapted to produce gene-single chain antibodies. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed (Huse, *et al., Science*, 246:1275-81 (1989)) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Screening Methods

The present invention may be employed in a process for screening for agents such as agonists, *i.e.* agents that bind to and activate target gene polypeptides, or antagonists, *i.e.* inhibit the activity or interaction of target gene polypeptides with its ligand. Thus, polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures as known in the art. Any methods routinely used to identify and screen for agents that can modulate receptors may be used in accordance with the present invention.

The present invention provides methods for identifying and screening for agents that modulate target gene expression or function. More particularly, cells that contain and express target gene sequences may be used to screen for therapeutic agents. Such cells may include non-recombinant monocyte cell lines, such as U937 (ATCC# CRL-1593), THP-1 (ATCC# TIB-202), and P388D1 (ATCC# TIB-63); endothelial cells such as HUVEC's and bovine aortic endothelial cells (BAEC's); as well as generic mammalian cell lines such as HeLa cells and COS cells, *e.g.*, COS-7 (ATCC# CRL-1651). Further, such cells may include recombinant, transgenic cell lines. For example, the transgenic mice of the invention may be used to generate cell lines, containing one or

more cell types involved in a disease, that can be used as cell culture models for that disorder. While cells, tissues, and primary cultures derived from the disease transgenic animals of the invention may be utilized, the generation of continuous cell lines is preferred. For examples of techniques which may be used to derive a continuous cell line from the transgenic animals, *see* Small, *et al.*, *Mol. Cell Biol.*, 5:642-48 (1985).

Target gene sequences may be introduced into, and overexpressed in, the genome of the cell of interest. In order to overexpress a target gene sequence, the coding portion of the target gene sequence may be ligated to a regulatory sequence which is capable of driving gene expression in the cell type of interest. Such regulatory regions will be well known to those of skill in the art, and may be utilized in the absence of undue experimentation. Target gene sequences may also be disrupted or underexpressed. Cells having target gene disruptions or underexpressed target gene sequences may be used, for example, to screen for agents capable of affecting alternative pathways which compensate for any loss of function attributable to the disruption or underexpression.

In vitro systems may be designed to identify compounds capable of binding the target gene products. Such compounds may include, but are not limited to, peptides made of D-and/or L-configuration amino acids (in, for example, the form of random peptide libraries; see e.g., Lam, *et al.*, *Nature*, 354:82-4 (1991)), phosphopeptides (in, for example, the form of random or partially degenerate, directed phosphopeptide libraries; see, e.g., Songyang, *et al.*, *Cell*, 72:767-78 (1993)), antibodies, and small organic or inorganic molecules. Compounds identified may be useful, for example, in modulating the activity of target gene proteins, preferably mutant target gene proteins; elaborating the biological function of the target gene protein; or screening for compounds that disrupt normal target gene interactions or themselves disrupt such interactions.

The principle of the assays used to identify compounds that bind to the target gene protein involves preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex which can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring the target gene protein or the test substance onto a solid phase and detecting target protein/test substance complexes anchored on the solid phase at the end of the reaction. In one embodiment of such a method, the target gene protein may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly.

In practice, microtitre plates are conveniently utilized. The anchored component may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished simply by coating the solid surface with a solution of the protein and drying. Alternatively, an immobilized antibody, preferably a monoclonal antibody, specific for the protein may be used to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the nonimmobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously nonimmobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously nonimmobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the previously nonimmobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

Alternatively, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for target gene product or the test compound to anchor any complexes formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes.

Compounds that are shown to bind to a particular target gene product through one of the methods described above can be further tested for their ability to elicit a biochemical response from the target gene protein. Agonists, antagonists and/or inhibitors of the expression product can be identified utilizing assays well known in the art.

Antisense, Ribozymes, and Antibodies

Other agents which may be used as therapeutics include the target gene, its expression product(s) and functional fragments thereof. Additionally, agents which reduce or inhibit mutant target gene activity may be used to ameliorate disease symptoms. Such agents include antisense, ribozyme, and triple helix molecules. Techniques for the production and use of such molecules are well known to those of skill in the art.

Anti-sense RNA and DNA molecules act to directly block the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest, are preferred.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by an endonucleolytic cleavage. The composition of ribozyme molecules must include one or more sequences complementary to the target gene mRNA, and must include the well known catalytic sequence responsible for mRNA cleavage. For this sequence, see U.S. Pat. No. 5,093,246, which is incorporated by reference herein in its entirety. As such within the scope of the invention are engineered hammerhead motif ribozyme molecules that

specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences encoding target gene proteins.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the molecule of interest for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features, such as secondary structure, that may render the oligonucleotide sequence unsuitable. The suitability of candidate sequences may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

Nucleic acid molecules to be used in triple helix formation for the inhibition of transcription should be single stranded and composed of deoxyribonucleotides. The base composition of these oligonucleotides must be designed to promote triple helix formation via Hoogsteen base pairing rules, which generally require sizeable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences may be pyrimidine-based, which will result in TAT and CGC- triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide base complementarity to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules may be chosen that are purine-rich, for example, containing a stretch of G residues. These molecules will form a triple helix with a DNA duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in GGC triplets across the three strands in the triplex.

Alternatively, the potential sequences that can be targeted for triple helix formation may be increased by creating a so called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

It is possible that the antisense, ribozyme, and/or triple helix molecules described herein may reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by both normal and mutant target gene alleles. In order to ensure that substantially normal levels of target gene activity are maintained, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal activity may be introduced into cells that do not contain sequences susceptible to whatever antisense, ribozyme, or triple helix treatments are being utilized. Alternatively, it may be preferable to coadminister normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

Anti-sense RNA and DNA, ribozyme, and triple helix molecules of the invention may be prepared by any method known in the art for the synthesis of DNA and RNA molecules. These

include techniques for chemically synthesizing oligodeoxyribonucleotides and oligoribonucleotides well known in the art such as for example solid phase phosphoramidite chemical synthesis. Alternatively, RNÁ molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Various well-known modifications to the DNA molecules may be introduced as a means of increasing intracellular stability and half-life. Possible modifications include but are not limited to the addition of flanking sequences of ribonucleotides or deoxyribonucleotides to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the oligodeoxyribonucleotide backbone.

Antibodies that are both specific for target gene protein, and in particular, mutant gene protein, and interfere with its activity may be used to inhibit mutant target gene function. Such antibodies may be generated against the proteins themselves or against peptides corresponding to portions of the proteins using standard techniques known in the art and as also described herein. Such antibodies include but are not limited to polyclonal, monoclonal, Fab fragments, single chain antibodies, chimeric antibodies, etc.

In instances where the target gene protein is intracellular and whole antibodies are used, internalizing antibodies may be preferred. However, lipofectin liposomes may be used to deliver the antibody or a fragment of the Fab region which binds to the target gene epitope into cells. Where fragments of the antibody are used, the smallest inhibitory fragment which binds to the target or expanded target protein's binding domain is preferred. For example, peptides having an amino acid sequence corresponding to the domain of the variable region of the antibody that binds to the target gene protein may be used. Such peptides may be synthesized chemically or produced via recombinant DNA technology using methods well known in the art (see, *e.g.*, Creighton, Proteins: Structures and Molecular Principles (1984) W.H. Freeman, New York 1983, *supra*; and Sambrook, *et al.*, 1989, supra). Alternatively, single chain neutralizing antibodies which bind to intracellular target gene epitopes may also be administered. Such single chain antibodies may be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population by utilizing, for example, techniques such as those described in Marasco, *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:7889-93 (1993).

RNA sequences encoding a target gene protein may be directly administered to a patient exhibiting disease symptoms, at a concentration sufficient to produce a level of target gene protein such that disease symptoms are ameliorated. Patients may be treated by gene replacement therapy. One or more copies of a normal target gene, or a portion of the gene that directs the production of a

normal target gene protein with target gene function, may be inserted into cells using vectors which include, but are not limited to adenovirus, adeno-associated virus, and retrovirus vectors, in addition to other particles that introduce DNA into cells, such as liposomes. Additionally, techniques such as those described above may be utilized for the introduction of normal target gene sequences into human cells.

Cells, preferably, autologous cells, containing normal target gene expressing gene sequences may then be introduced or reintroduced into the patient at positions which allow for the amelioration of disease symptoms.

Pharmaceutical Compositions, Effective Dosages, and Routes of Administration

The identified compounds that inhibit target mutant gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to treat or ameliorate the disease. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disease.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral, topical, subcutaneous, intraperitoneal, intravenous, intrapleural, intraoccular,

intraarterial, or rectal administration. It is also contemplated that pharmaceutical compositions may be administered with other products that potentiate the activity of the compound and optionally, may include other therapeutic ingredients.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g. gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. Oral ingestion is possibly the easiest method of taking any medication. Such a route of administration, is generally simple and straightforward and is frequently the least inconvenient or unpleasant route of administration from the patient's point of view. However, this involves passing the material through the stomach, which is a hostile environment for many materials, including proteins and other biologically active compositions. As the acidic, hydrolytic and proteolytic environment of the stomach has evolved efficiently to digest proteinaceous materials into amino acids and oligopeptides for subsequent anabolism, it is hardly surprising that very little or any of a wide variety of biologically active proteinaceous material, if simply taken orally, would survive its passage through the stomach to be taken up by the body in the small intestine. The result, is that many proteinaceous medicaments must be taken in through another method, such as parenterally, often by subcutaneous, intramuscular or intravenous injection.

Pharmaceutical compositions may also include various buffers (*e.g.*, Tris, acetate, phosphate), solubilizers (*e.g.*, Tween, Polysorbate), carriers such as human serum albumin, preservatives (thimerosol, benzyl alcohol) and anti-oxidants such as ascorbic acid in order to stabilize pharmaceutical activity. The stabilizing agent may be a detergent, such as tween-20, tween-80, NP-40 or Triton X-100. EBP may also be incorporated into particulate preparations of polymeric compounds for controlled delivery to a patient over an extended period of time. A more extensive survey of components in pharmaceutical compositions is found in Remington's Pharmaceutical Sciences, 18th ed., A. R. Gennaro, ed., Mack Publishing, Easton, Pa. (1990).

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

Diagnostics

A variety of methods may be employed to diagnose disease conditions associated with the target gene. Specifically, reagents may be used, for example, for the detection of the presence of target gene mutations, or the detection of either over or under expression of target gene mRNA.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type target gene locus is detected. In addition, the method can be performed by detecting the wild-type target gene locus and confirming the lack of a predisposition or neoplasia. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, *e.g.*, in the tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are mutated, then a late neoplastic state may be indicated. The finding of gene mutations thus provides both diagnostic and prognostic information. A target gene allele which is not deleted (e.g., that found on the sister chromosome to a chromosome carrying a target gene deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. Mutations found in tumor tissues may be linked to decreased expression of the target gene product. However, mutations leading to non-functional gene products may also be linked to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the target gene product, or a decrease in mRNA stability or translation efficiency.

One test available for detecting mutations in a candidate locus is to directly compare genomic target sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, *e.g.*, by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene. Mutations from cancer patients falling outside the coding region of the target gene can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the target gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific gene nucleic acid or anti-gene antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings, to diagnose patients exhibiting disease symptoms or at risk for developing disease.

Any cell type or tissue, preferably monocytes, endothelial cells, or smooth muscle cells, in which the gene is expressed may be utilized in the diagnostics described below.

DNA or RNA from the cell type or tissue to be analyzed may easily be isolated using procedures which are well known to those in the art. Diagnostic procedures may also be performed *in situ* directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resec-

tions, such that no nucleic acid purification is necessary. Nucleic acid reagents may be used as probes and/or primers for such *in situ* procedures (see, for example, Nuovo, PCR *In Situ* Hybridization: Protocols and Applications, Raven Press, N.Y. (1992)).

Gene nucleotide sequences, either RNA or DNA, may, for example, be used in hybridization or amplification assays of biological samples to detect disease-related gene structures and expression. Such assays may include, but are not limited to, Southern or Northern analyses, restriction fragment length polymorphism assays, single stranded conformational polymorphism analyses, *in situ* hybridization assays, and polymerase chain reaction analyses. Such analyses may reveal both quantitative aspects of the expression pattern of the gene, and qualitative aspects of the gene expression and/or gene composition. That is, such aspects may include, for example, point mutations, insertions, deletions, chromosomal rearrangements, and/or activation or inactivation of gene expression.

Preferred diagnostic methods for the detection of gene-specific nucleic acid molecules may involve for example, contacting and incubating nucleic acids, derived from the cell type or tissue being analyzed, with one or more labeled nucleic acid reagents under conditions favorable for the specific annealing of these reagents to their complementary sequences within the nucleic acid molecule of interest. Preferably, the lengths of these nucleic acid reagents are at least 9 to 30 nucleotides. After incubation, all non-annealed nucleic acids are removed from the nucleic acid:fingerprint molecule hybrid. The presence of nucleic acids from the fingerprint tissue which have hybridized, if any such molecules exist, is then detected. Using such a detection scheme, the nucleic acid from the tissue or cell type of interest may be immobilized, for example, to a solid support such as a membrane, or a plastic surface such as that on a microtitre plate or polystyrene beads. In this case, after incubation, non-annealed, labeled nucleic acid reagents are easily removed. Detection of the remaining, annealed, labeled nucleic acid reagents is accomplished using standard techniques well-known to those in the art.

Alternative diagnostic methods for the detection of gene-specific nucleic acid molecules may involve their amplification, e.g., by PCR (the experimental embodiment set forth in Mullis U.S. Pat. No. 4,683,202 (1987)), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. USA*, 88:189-93 (1991)), self sustained sequence replication (Guatelli, *et al.*, *Proc. Natl. Acad. Sci.* USA, 87:1874-78 (1990)), transcriptional amplification system (Kwoh, *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:1173-77 (1989)), Q-Beta Replicase (Lizardi *et al.*, *Bio/Technology*, 6:1197 (1988)), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In one embodiment of such a detection scheme, a cDNA molecule is obtained from an RNA molecule of interest (*e.g.*, by reverse transcription of the RNA molecule into cDNA). Cell types or

tissues from which such RNA may be isolated include any tissue in which wild type fingerprint gene is known to be expressed, including, but not limited, to monocytes, endothelium, and/or smooth muscle. A sequence within the cDNA is then used as the template for a nucleic acid amplification reaction, such as a PCR amplification reaction, or the like. The nucleic acid reagents used as synthesis initiation reagents (*e.g.*, primers) in the reverse transcription and nucleic acid amplification steps of this method may be chosen from among the gene nucleic acid reagents described herein. The preferred lengths of such nucleic acid reagents are at least 15-30 nucleotides. For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining or by utilizing any other suitable nucleic acid staining method.

Antibodies directed against wild type or mutant gene peptides may also be used as disease diagnostics and prognostics. Such diagnostic methods, may be used to detect abnormalities in the level of gene protein expression, or abnormalities in the structure and/or tissue, cellular, or subcellular location of fingerprint gene protein. Structural differences may include, for example, differences in the size, electronegativity, or antigenicity of the mutant fingerprint gene protein relative to the normal fingerprint gene protein.

Protein from the tissue or cell type to be analyzed may easily be detected or isolated using techniques which are well known to those of skill in the art, including but not limited to western blot analysis. For a detailed explanation of methods for carrying out western blot analysis. (*See, e.g.* Sambrook, *et al.* (1989) *supra*, at Chapter 18). The protein detection and isolation methods employed herein may also be such as those described in Harlow and Lane, for example, (Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988)).

Preferred diagnostic methods for the detection of wild type or mutant gene peptide molecules may involve, for example, immunoassays wherein fingerprint gene peptides are detected by their interaction with an anti-fingerprint gene-specific peptide antibody.

For example, antibodies, or fragments of antibodies useful in the present invention may be used to quantitatively or qualitatively detect the presence of wild type or mutant gene peptides. This can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow cytometric, or fluorimetric detection. Such techniques are especially preferred if the fingerprint gene peptides are expressed on the cell surface.

The antibodies (or fragments thereof) useful in the present invention may, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of fingerprint gene peptides. *In situ* detection may be accomplished by removing a histological specimen from a patient, and applying thereto a labeled antibody of the present invention.

The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the fingerprint gene peptides, but also their distribution in the examined tissue. Using the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

Immunoassays for wild type, mutant, or expanded fingerprint gene peptides typically comprise incubating a biological sample, such as a biological fluid, a tissue extract, freshly harvested cells, or cells which have been incubated in tissue culture, in the presence of a detectably labeled antibody capable of identifying fingerprint gene peptides, and detecting the bound antibody by any of a number of techniques well known in the art.

The biological sample may be brought in contact with and immobilized onto a solid phase support or carrier such as nitrocellulose, or other solid support which is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled gene-specific antibody. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on solid support may then be detected by conventional means.

The terms "solid phase support or carrier" are intended to encompass any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

The binding activity of a given lot of anti-wild type or -mutant fingerprint gene peptide antibody may be determined according to well known methods. Those skilled in the art will be able to determine operative and optimal assay conditions for each determination by employing routine experimentation.

One of the ways in which the gene peptide-specific antibody can be detectably labeled is by linking the same to an enzyme and using it in an enzyme immunoassay (EIA) (Voller, *Ric Clin Lab*, 8:289-98 (1978) ["The Enzyme Linked Immunosorbent Assay (ELISA)", Diagnostic Horizons 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, Md.]; Voller, *et al.*, *J. Clin.*

Pathol., 31:507-20 (1978); Butler, *Meth. Enzymol.*, 73:482-523 (1981); Maggio (ed.), Enzyme Immunoassay, CRC Press, Boca Raton, Fla. (1980); Ishikawa, *et al.*, (eds.) Enzyme Immunoassay, Igaku-Shoin, Tokyo (1981)). The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect fingerprint gene wild type, mutant, or expanded peptides through the use of a radioimmunoassay (RIA) (see, *e.g.*, Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as 152Eu, or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediamine-tetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in, which a

catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Throughout this application, various publications, patents and published patent applications are referred to by an identifying citation. The disclosures of these publications, patents and published patent specifications referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

The following examples are intended only to illustrate the present invention and should in no way be construed as limiting the subject invention.

Examples

Example 1: Generation and Analysis of Mice Comprising Anaphylatoxin C3a Receptor Gene Disruptions

To investigate the role of anaphylatoxin C3a receptors, disruptions in anaphylatoxin C3a receptor genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in anaphylatoxin C3a receptor genes were created. More particularly, as shown in Figure 2A-2B, an anaphylatoxin C3a receptor-specific targeting construct having the ability to disrupt or modify anaphylatoxin C3a receptor genes, specifically comprising SEQ ID NO:1 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:3 or SEQ ID NO:4.

The targeting construct was introduced into ES cells derived from the 129/Sv+P+Mgf-SLJ/J mouse substrain to generate chimeric mice. The F1 mice were generated by breeding with C57BL/6 females, and the F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

The transgenic mice comprising disruptions in anaphylatoxin C3a receptor genes were analyzed for phenotypic changes and expression patterns. The phenotypes associated with a disruption in nuclear receptor genes were determined. The homozygous mice demonstrated at least one of the following phenotypes:

Thymus. Abnormalities in the thymus, including reduced size and reduced weight of the thymus in the mutant mice as compared to wild-type mice. Specifically, homozygous mice were reported to have a small thymus at necropsy as well as reduced thymus weights and reduced thymus to body weight ratio as compared to wild-type mice as shown in Figure 3 and the following Table 1:

Anaphylatoxin C3a Receptor

	0.53				Thymus	Thymus/
	Mouse▲	Age at Test	Length	Body Weight	Weight	Body Weight
		(days)	(cm)	(g)	(%)	(g)
+/+	Female					
	37302	52	9	19.22	0.07	0.38
	43019	49	10	19.4	0.08	0.4
+/+	Male					
	37294	49	10.5	23.95	0.08	0.32
	43015	49	9.5	26.22	0.05	0.19
-/+	Female					
	37301	52	10	21.14	0.09	0.41
-/+	Male					
	37307	49	10.25	25.63	0.07	0.28
-/-	Female					
	37303	52	9.75	19.72	0.08	0.38
	43021	49	9	18.89	0.1	0.53
	43022	49	9	22.18	0.09	0.4
-/-	Male					
	37306	49	9	25.78	0.03	0.13
	43016	49	10.5	24.64	0.04	0.16
	43018	49	10.25	27.46	0.06	0.2

TABLE 1

Expression. Total RNA was isolated from the organs or tissues from adult C57Bl/6 wild type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA transcripts were detectable in are detectable in brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, spinal cord, eye, Harderian gland, heart, lung, liver, pancreas, kidney, spleen, thymus, lymph nodes, bone marrow, skin, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovary, uterus and white fat. No RNA transcripts were detectable in gallbladder.

Behavior:

For behavioral studies, homozygous mice were produced as follows:

The targeting construct described above was introduced into ES cells derived from the 129/SvEv mouse substrain to generate chimeric mice. F1N0 mice were generated by breeding with C57BL/6 females. F2N0 homozygous mutant mice were produced by intercrossing F1 heterozygous

males and females. F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mice were produced by intercrossing F1N1 heterozygous males and females.

The homozygous mice demonstrated the following behavioral phenotypes:

Homozygous mice required a significantly smaller dose of metrazol to elicit characteristic seizure-like responses, a indicating an increased susceptibility to seizure. Homozygous mutants also displayed significantly decreased PPI, with a 90dB prepulse, indicating a stimulus processing deficit similar to that observed in schizophrenic patients.

Example 2: Generation and Analysis of Mice Comprising 5-HT5A Gene Disruptions

To investigate the role of 5-HT5A genes, disruptions in 5-HT5A genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in 5-HT5A genes were created. More particularly, as shown in Figure 5A-5B, an 5-HT5A-specific targeting construct having the ability to disrupt or modify 5-HT5A genes, specifically comprising SEQ ID NO:5 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:7 or SEQ ID NO:8.

The targeting construct was introduced into ES cells derived from the 129/Sv-+P+Mgf-SLJ/J mouse substrain to generate chimeric mice. The F1 mice were generated by breeding with C57BL/6 females, and the F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

The transgenic mice comprising disruptions in 5-HT5A genes were analyzed for phenotypic changes and expression patterns.

Expression. LacZ (beta-galactosidase) expression was detectable in the brain and esophagus. In the brain, staining was restricted to the cerebellum with Purkinje cells and nuclei in the granular layer displaying expression. In wholemount staining, lacZ expression was detectable solely in the cerebellum. Coronal sections of the cerebellum revealed that X-Gal signals were restricted to the Purkinje cell layer. In addition, distinct nuclei adjacent to the Purkinje cells exhibit staining. Strong X-gal signals were also present in squamous epithelial cells.

Example 3: Generation and Analysis of Mice Comprising Chordin Gene Disruptions

To investigate the role of chordins, disruptions in chordin genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in chordin genes were created. More particularly, as shown in Figure 5A-5B, a chordin-specific targeting construct having the ability to disrupt or modify chordin genes, specifically comprising SEQ ID NO:9 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:11 or SEQ ID NO:12.

The targeting construct was introduced into ES cells derived from the 129/Sv-+P+Mgf-SLJ/J mouse substrain to generate chimeric mice. The F1 mice were generated by breeding with C57BL/6

females, and the F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

The transgenic mice comprising disruptions in chordin genes were analyzed for phenotypic changes and expression patterns. The phenotypes associated with a disruption in chordin genes were determined. The homozygous mice demonstrated at least one of the following phenotypes: abnormal nociception, decreased response to pain and anxiety.

Expression. Total RNA was isolated from the organs or tissues from adult C57Bl/6 wild type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA transcripts were detectable in brain, cortex, subcortical region, cerebellum, brainstem, eye, heart, lung, liver, pancreas, kidneys, skin, gallbladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, tongue, stomach, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovary and uterus.

Behavior:

For behavioral studies, homozygous mice were produced as follows:

The targeting construct described above was introduced into ES cells derived from the 129/SvEv mouse substrain to generate chimeric mice. F1N0 mice were generated by breeding with C57BL/6 females. F2N0 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females. F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mice were produced by intercrossing F1N1 heterozygous males and females.

The homozygous mice demonstrated the following behavioral phenotypes:

F2N0 Homozygous mutant mice showed a statistically significant increase in their response latency to the hot plate test, relative to wild-type animals. The homozygous mice showed a statistically significant increase in the amount of time to lick their hindpaw when placed on the hot plate at 55°C, compared to F2N0 wild type mice. However, N1 generation animals did not display any difference. This test has been used to indicate a nociceptive disorder, *i.e.,* decreased pain response.

Homozygous mutant animals from both the N0 and N1 generation displayed a trend towards spending more time in the central region of the open field test chamber. This test may indicate an anxiety disorder.

Example 4: Generation and Analysis of Mice Comprising RORγ Gene Disruptions

To investigate the role of RORγs, disruptions in RORγ genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in RORγ genes were created. More particularly, as shown in Figure 9A-9B, a RORγ-specific targeting construct having the ability to disrupt or modify RORγ genes, specifically comprising SEQ ID NO:13 was created using as the

targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:15 or SEQ ID NO:16.

The targeting construct was introduced into ES cells derived from the 129/Sv-+P+Mgf-SLJ/J mouse substrain to generate chimeric mice. The F1 mice were generated by breeding with C57BL/6 females, and the F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

The transgenic mice comprising disruptions in RORγ genes were analyzed for phenotypic changes and expression patterns. The phenotypes associated with a disruption in nuclear receptor genes were determined. The homozygous mice demonstrated at least one of the following phenotypes:

Lymph Node. Lymph nodes were absent in all female homozygous mice. Furthermore, gut associated lymphoid tissue (GALT) was absent in routine sections of intestine from all homozygous mutants.

Lymphocytes. Minimal to mild accumulations of lymphocytes were present in a variety of tissues (liver, lung, pancreas, salivary gland and thyroid gland) in mutant mice as compared to wild-type mice. Homozygous mutant mice had multiple organs with lymphoid infiltrates consistent with lymphoma. The infiltrates varied from mild-to-severe and can involve any organ system. These cells had an increased mitotic rate and frequent apoptosis. Involvement tends to be most severe in the spleen and thymus. Other organs involved include: liver, gallbladder, lung, kidneys, urinary bladder, heart, aorta, bones, bone marrow, pituitary gland, adrenal glands, thyroid gland, brain and spinal cord (meninges), pancreas, stomach, small and large intestines, larynx, trachea, esophagus, tongue, skeletal muscle, testes, epididymides, urinary bladder, seminal vesicles, eyes, and Harderian glands.

2 clinically ill homozygous females (20774 and 20299) and 2 clinically ill homozygous males (19922 and 20294) had greatly elevated total white blood cell counts in which lymphocytes predominated. In one leukemic female 20774, lymphocytes were immature (blastic). This is consistent with a leukemic phase of lymphoma. Affected mice homozygous showed signs of illness, including poor grooming and hypoactivity. The mice had increased thoracic and abdominal girth, hepatosplenomegaly, enlarged thymuses, and ascites; they appeared weak, squint, are hunched, cachectic, lethargic, and have impaired respiration.

Spleen. Abnormalities in the spleen were detected, including increased weight and increased spleen:body weight ratios in all homozygous mutant relative to wild-type mice, as shown in Table 2 and Figure 10, below. These averaged 0.4 percent in wild-type control females vs. 0.61 percent in homozygous females, and averaged 0.26 percent in wild-type control males and 0.52 percent in homozygous mutant males. At 68 - 189 days spleen weights and spleen:body weight ratios were approximately 5 times greater than in wild-type mice.

Splenic lymphoid follicles were expanded by increased numbers of lymphocytes (hyperplasia) in germinal centers and at the marginal zone of the follicles in homozygous mice.

RORγ (Orphan NHR)

					Spleen	Spleen/	Liver	Liver/	Kidney	Kidney/	Thymus	Thymus/
	Mouse▲	Age at Test	Length	Body Weight	Weight	Body Weight	Weight	Body Weight	Weight	Body Weight	Weight	Body Weight
		(days)	(cm)	(g)		(%)	(g)	(%)	(g)	(%)	(g)	(%)
+/+	Female											
	20139	313	9.5	24.66	0.17	0.67	1.19	4.81	0.36	1.46	0.04	0.15
	20140	313	10.2	26.59	0.1	0.36	0.98	3.69	0.28	1.05	0.07	0.26
+/+	Male											
	19319	321	10.6	39.94	0.1	0.26	1.94	4.84	0.61	1.53	0.04	0.09
	19917	321	10	33.23	0.08	0.24	1.51	4.55	0.57	1.72	0.05	0.15
-/-	Female											
	20141	313	10.4	30.54	0.15	0.47	1.11	3.63	0.34	1.12	0.07	0.24
	20839	312	10.1	25.57	0.35	1.37	1.52	5.93	0.42	1.63	0.06	0.25
-/-	Male											
	19918	321	10.5	34.63	0.22	0.64	1.65	4.76	0.56	1.6	0.04	0.11
	19923	321	11	36.18	0.18	0.48	1.57	4.34	0.59	1.62	0.04	0.1

TABLE 2

Liver. Abnormalities in the liver were detected, including increased size, weight, and discoloration in mutant mice as compared to wild-type mice. Liver weight and liver:body weight ratios were approximately 2 times greater than in wild-type control mice (Table 2 and Figure 11).

Kidney. Abnormalities in the kidney were detected, including increased size, weight and discoloration in mutant mice as compared to wild-type mice. Kidney:body weight ratios were approximately 1.5 times greater than in wild-type control mice (Table 2 and Figure 12).

Thymus. Abnormalities in the thymus were detected. Thymus weight and thymus:body weight ratios were approximately 10 to 30 times greater than in wild-type control mice (Table 1 and Figure 13). There was moderate thymic atrophy with loss of corticomedullary distinction. In addition, thymic cortical expansion (hyperplasia) and associated medullary reduction (atrophy) was observed in all homozygous mice. Thymic cortical lymphocytes were hyperplastic with increased mitotic activity and apoptosis.

Bone and Bone Marrow. Abnormalities in the bones and bone marrow were detected at 68 - 189 days. Bone marrow was pale and bones were brittle with white masses attached.

Body Weight. Male homozygous mice body weights were approximately 60% those of wild-type control mice. There was no difference in body weight for female homozygous mutant mice versus wild-type control mice. Gross observations were incomplete for some mice.

Serum Chemistry. Clinically ill homozygous mice had elevations of various analytes when compared to wild-type mice. Multiple liver-related analytes were elevated, including: alanine aminotransferase (ALT) in all 4 males and in 2 of 3 females: bilirubin in 2 males; and, alkaline phosphatase (ALP) in one male. Aspartate aminotransferase, another liver-related analyte, was elevated in 4 males and 2 females. These various elevations are consistent with tumor cell infiltration

of the liver. Tumor cell infiltration of the heart could also have contributed to ALT elevations. Kidney-related analytes were also elevated. Three males had elevated blood urea nitrogen (BUN) and phosphorus levels, possibly related to tumor cell infiltration of the kidneys and/or renal pelvics.

Expression. Total RNA was isolated from the organs or tissues from adult C57Bl/6 wild type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA transcripts were detectable in cerebellum, olfactory bulb, eye, heart, lung, liver, pancreas, kidneys, thymus, lymph nodes, skin, gallbladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovary and uterus.

Example 5: Generation and Analysis of Mice Comprising BMP Gene Disruptions

To investigate the role of BMP genes, disruptions in BMP genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in BMP genes were created. More particularly, as shown in Figure 15A-15B, a BMP-specific targeting construct having the ability to disrupt or modify BMP genes, specifically comprising SEQ ID NO:17 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:19 or SEQ ID NO:20.

The targeting construct was introduced into ES cells derived from the 129/Sv-+P+Mgf-SLJ/J mouse substrain to generate chimeric mice. The F1 mice were generated by breeding with C57BL/6 females, and the F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

The transgenic mice comprising disruptions in BMP genes were analyzed for phenotypic changes. The phenotypes associated with a disruption in nuclear receptor genes were determined. The homozygous mice demonstrated at least one of the following phenotypes:

1) Kinky tail:

   At 49 days, when compared to age- and gender-matched wild-type control mice, 5 of the 9 homozygous mice examined, 2 females (10796, 10799) and 3 males (10829, 19478, 19479), had kinky tails.

   At 300 days (actual ages 303 to 450 days), when compared to age- and gender-matched wild-type control mice, 1 male (10827) of the 9 homozygous mutant mice had a kinky tail.

2) Body Weight:

   Low body weight trend, homozygous mutant males and females at 49, 90, 180, and 300 days of age.

Low body weight / body length ratio trend, homozygous mutant males and females at 49, 90, 180, and 300 days of age.

3) Short Body Length:

Homozyogous mutant males and females at 49, 90, 180, and 300 days of age.

More specifically, when compared to age- (49, 90, 180, and 300 days) and gender-matched wild-type control mice, the homozygous mutant mice had lower values for all 3 parameters (body weight, body length, and body weight/body length ratio) at all time points. Generally the differences between homozygous mutant and matched wild-type control mice were greater for the males than for the females. The body weight and the body weight/body length ratio decreased for homozygous mutant males after 180 days of age. For each of the 3 parameters, the magnitude of the difference between wild-type control and homozygous mutant males was greatest at 300 days of age.

Example 6: Generation and Analysis of Mice Comprising Airway Trypsin-like Protease Gene Disruptions

To investigate the role of airway trypsin-like proteases, disruptions in airway trypsin-like protease genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in airway trypsin-like protease genes were created. More particularly, as shown in Figure 17A-17B, a airway trypsin-like protease-specific targeting construct having the ability to disrupt or modify airway trypsin-like protease genes, specifically comprising SEQ ID NO:21 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:22 or SEQ ID NO:24.

The targeting construct was introduced into ES cells derived from the 129/Sv-+P+Mgf-SLJ/J mouse substrain to generate chimeric mice. The F1 mice were generated by breeding with C57BL/6 females, and the F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females. The transgenic mice comprising disruptions in airway trypsin-like protease genes were analyzed for phenotypic changes.

Example 7: Generation and Analysis of Mice Comprising Aquaporin-8 Gene Disruptions

To investigate the role of aquaporin-8 genes, disruptions in aquaporin-8 genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in aquaporin-8 genes were created. More particularly, as shown in Figure 19A-19B, an aquaporin-8 -specific targeting construct having the ability to disrupt or modify aquaporin-8 genes, specifically comprising SEQ ID NO:25 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:27 or SEQ ID NO:28.

The targeting construct was introduced into ES cells derived from the 129/Sv-+P+Mgf-SLJ/J mouse substrain to generate chimeric mice. The F1 mice were generated by breeding with C57BL/6 females, and the F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

The transgenic mice comprising disruptions in aquaporin-8 genes were analyzed for phenotypic changes and expression patterns. The phenotypes associated with a disruption in nuclear receptor genes were determined. The homozygous mice demonstrated at least one of the following phenotypes:

Weight. Homozygous mutant males have a tendency to weigh more than wild-type control mice after age 90 days.

Expression. Total RNA was isolated from the organs or tissues from adult C57Bl/6 wild type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA transcripts were detectable in eye, liver, pancreas, thymus, gallbladder, stomach, large intestine, cecum, testis, seminal vesicle, ovary and uterus.

As is apparent to one of skill in the art, various modifications of the above embodiments can be made without departing from the spirit and scope of this invention. These modifications and variations are within the scope of this invention.

We claim:

1. A targeting construct comprising:
   (a) a first polynucleotide sequence homologous to a target gene;
   (b) a second polynucleotide sequence homologous to the target gene; and
   (c) a selectable marker.

2. The targeting construct of claim 1, wherein the targeting construct further comprises a screening marker.

3. A method of producing a targeting construct for a target gene, the method comprising:
   (a) obtaining a first polynucleotide sequence homologous to a target gene;
   (b) obtaining a second polynucleotide sequence homologous to the target gene;
   (c) providing a vector comprising a selectable marker; and
   (d) inserting the first and second sequences into the vector, to produce the targeting construct.

4. A method of producing a targeting construct for a target gene, the method comprising:
   (a) providing a polynucleotide comprising a first sequence homologous to a first region of a target gene and a second sequence homologous to a second region of a target gene; and
   (b) inserting a positive selection marker in between the first and second sequences to form the targeting construct.

5. A cell comprising a disruption in a target gene.

6. The cell of claim 5, wherein the cell is a murine cell.

7. The cell of claim 6, wherein the murine cell is an embryonic stem cell.

8. A non-human transgenic animal comprising a disruption in a target gene.

9. A cell derived from the non-human transgenic animal of claim 8.

10. A method of producing a transgenic mouse comprising a disruption in a target gene, the method comprising:
   (a) introducing the targeting construct of claim 1 into a cell;
   (b) introducing the cell into a blastocyst;
   (c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and
   (d) breeding the chimeric mouse to produce the transgenic mouse.

11. A method of identifying an agent that modulates the expression of a target gene, the method comprising:
   (a) providing a non-human transgenic animal comprising a disruption in the target gene;
   (b) administering an agent to the non-human transgenic animal; and
   (c) determining whether the expression of the disrupted target gene in the non-human transgenic animal is modulated.

12. A method of identifying an agent that modulates the function of a target gene, the method comprising:

(a) providing a non-human transgenic animal comprising a disruption in the target gene;

(b) administering an agent to the non-human transgenic animal; and

(c) determining whether the function of the disrupted target gene in the non-human transgenic animal is modulated.

13. A method of identifying an agent that modulates the expression of a target gene, the method comprising:

(a) providing a cell comprising a disruption in a target gene;

(b) contacting the cell with an agent; and

(c) determining whether expression of the target gene is modulated.

14. A method of identifying an agent that modulates the function of a target gene, the method comprising:

(a) providing a cell comprising a disruption in a target gene;

(b) contacting the cell with an agent; and

(c) determining whether the function of the target gene is modulated.

15. The method of claim 13 or claim 14, wherein the cell is derived from the non-human transgenic animal of claim 8.

16. An agent identified by the method of claim 11, claim 12, claim 13, or claim 14.

17. An agent that modulates the activity of a target gene.

18. A transgenic mouse comprising a disruption in an anaphylatoxin C3a receptor gene, wherein the transgenic mouse exhibits an abnormality of the thymus.

19. The transgenic mouse of claim 18, wherein the thymus abnormality is reduced weight of the thymus relative to a wild-type mouse.

20. The transgenic mouse of claim 18, wherein the thymus abnormality is a reduced thymus to body weight ratio relative to a wild-type mouse.

21. A transgenic mouse comprising a disruption in an anaphylatoxin C3a receptor gene, wherein the transgenic mouse exhibits an increased susceptibility to seizure.

22. The transgenic mouse of claim 21, wherein the mouse exhibits seizure-like responses at a lower dose of Metrazol relative to a wild-type mouse.

23. A transgenic mouse comprising a disruption in an anaphylatoxin C3a receptor gene, wherein the transgenic mouse exhibits a stimulus processing deficit relative to a wild-type mouse.

24. The transgenic mouse of claim 23, wherein the stimulus processing deficit is similar to that observed in schizophrenia.

25. The transgenic mouse of claim 23, wherein the mouse exhibits decreased prepulse inhibition relative to a wild-type mouse.

26. A method of producing a transgenic mouse comprising a disruption in an anaphylatoxin C3a receptor gene, wherein the transgenic mouse exhibits at least one of the following phenotypes: an abnormality of the thymus, an increased susceptibility to seizure, or a stimulus processing deficit, the method comprising:

(a) introducing a anaphylatoxin C3a receptor gene targeting construct into a cell;

(b) introducing the cell into a blastocyst;

(c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

(d) breeding the chimeric mouse to produce the transgenic mouse comprising a disruption in an anaphylatoxin C3a receptor gene.

27. A cell derived from the transgenic mouse of claim 18 or claim 26.

28. A method of identifying an agent that ameliorates a phenotype associated with a disruption in an anaphylatoxin C3a receptor gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a anaphylatoxin C3a receptor gene; and

(b) determining whether the agent ameliorates at least one of the following phenotypes: an abnormality of the thymus, an increased susceptibility to seizure, or a stimulus processing deficit.

29. A method of identifying an agent that modulates anaphylatoxin C3a receptor expression, the method comprising:

(a) administering an agent to the transgenic mouse comprising a disruption in an anaphylatoxin C3a receptor gene; and

(b) determining whether the agent modulates anaphylatoxin C3a receptor expression in the transgenic mouse, wherein the agent has an effect on at least one of the following behaviors: susceptibility to seizure and stimulus processing.

30. A method of identifying an agent that modulates a behavior associated with a disruption in an anaphylatoxin C3a receptor gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in an anaphylatoxin C3a receptor gene; and

(b) determining whether the agent modulates at least one of the following behaviors: susceptibility to seizure and stimulus processing.

31. A method of identifying an agent that modulates anaphylatoxin C3a receptor gene function, the method comprising:

(a) providing a cell comprising a disruption in an anaphylatoxin C3a receptor gene;

(b) contacting the cell with an agent; and

(c) determining whether the agent modulates anaphylatoxin C3a receptor gene function, wherein the agent modulates a phenotype associated with a disruption in an anaphylatoxin C3a receptor gene.

32. The method of claim 31, wherein the phenotype comprises at least one of the following an abnormality of the thymus, an increased susceptibility to seizure, and a stimulus processing deficit.

33. An agent identified by the method of claim 28, claim 29, claim 30, or claim 31.

34. An agent that modulates the activity of an anaphylatoxin C3a receptor gene.

35. A method of ameliorating a disease associated with a mutation of an anaphylatoxin C3a receptor gene, the method comprising administering to a subject in need, a therapeutically effective amount of an agent that modulates the activity of an anaphylatoxin C3a receptor gene.

36. A transgenic mouse comprising a disruption in a chordin gene, wherein the transgenic mouse exhibits at least one of the following phenotypes: nociceptive disorder disorder and anxiety disorder.

37. A method of producing a transgenic mouse comprising a disruption in a chordin gene, wherein the transgenic mouse exhibits at least one of the following phenotypes: nociceptive disorder and anxiety disorder, the method comprising:

(a) introducing a chordin gene targeting construct into a cell;

(b) introducing the cell into a blastocyst;

(c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

(d) breeding the chimeric mouse to produce the transgenic mouse comprising a disruption in a chordin gene.

38. A cell derived from the transgenic mouse of claim 36 or claim 37.

39. A method of identifying an agent that ameliorates a phenotype associated with a disruption in a chordin gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a chordin gene; and

(b) determining whether the agent ameliorates at least one of the following phenotypes: nociceptive disorder and anxiety disorder.

40. A method of identifying an agent that modulates chordin expression, the method comprising:

(a) administering an agent to the transgenic mouse comprising a disruption in a chordin gene; and

(b) determining whether the agent modulates chordin expression in the transgenic mouse, wherein the agent has an effect on at least one of the following behaviors: response to pain and anxiety.

41. A method of identifying an agent that modulates a behavior associated with a disruption in a chordin gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a chordin gene; and

(b) determining whether the agent modulates coordination and balance of the transgenic mouse.

42. A method of identifying an agent that modulates chordin gene function, the method comprising:

(a) providing a cell comprising a disruption in a chordin gene;

(b) contacting the cell with an agent; and

(c) determining whether the agent modulates chordin gene function, wherein the agent modulates a phenotype associated with a disruption in a chordin gene.

43. The method of claim 42, wherein the phenotype comprises at least one of the following: nociceptive disorder and anxiety disorder.

44. An agent identified by the method of claim 39, claim 40, claim 41, or claim 42.

45. A transgenic mouse comprising a disruption in a chordin gene, wherein the transgenic mouse exhibits decreased response to pain.

46. An agent that modulates the activity of a chordin gene.

47. A method of ameliorating a disease associated with a mutation of a chordin gene, the method comprising administering to a subject in need, a therapeutically effective amount of an agent that modulates the activity of a chordin gene.

48. A transgenic mouse comprising a disruption in a RORγ gene, wherein the transgenic mouse exhibits at least one of the following phenotypes: a spleen abnormality, a kidney abnormality a spleen abnormality a liver abnormality, an abnormality of the thymus, an abnormality in the lymph nodes, an abnormality in the lymphocytes, an abnormality in the bone marrow, or an abnormality in the bones.

49. The transgenic mouse of claim 48, wherein the spleen abnormality is increased weight of the spleen relative to a wild-type mouse.

50. The transgenic mouse of claim 48, wherein the spleen abnormality is increased size of the spleen relative to a wild-type mouse.

51. The transgenic mouse of claim 48, wherein the spleen abnormality is an increased spleen to body weight ratio relative to a wild-type mouse.

52. The transgenic mouse of claim 48, wherein the kidney abnormality is increased weight of the kidney relative to a wild-type mouse.

53. The transgenic mouse of claim 48, wherein the kidney abnormality is increased size of the kidney relative to a wild-type mouse.

54. The transgenic mouse of claim 48, wherein the kidney abnormality is an increased kidney to body weight ratio relative to a wild-type mouse.

55. The transgenic mouse of claim 48, wherein the liver abnormality is increased weight of the liver relative to a wild-type mouse.

56. The transgenic mouse of claim 48, wherein the liver abnormality is increased size of the liver relative to a wild-type mouse.

57. The transgenic mouse of claim 48, wherein liver abnormality is an increased liver to body weight ratio relative to a wild-type mouse.

58. The transgenic mouse of claim 48, wherein the thymus abnormality is increased weight of the thymus relative to a wild-type mouse.

59. The transgenic mouse of claim 48, wherein the thymus abnormality is increased size of the thymus relative to a wild-type mouse.

60. The transgenic mouse of claim 48, wherein the thymus abnormality is an increased thymus to body weight ratio relative to a wild-type mouse.

61. The transgenic mouse of claim 48, wherein the abnormality of the thymus is thymic cortical expansion and medullary reduction relative to a wild-type mouse.

62. The transgenic mouse of claim 48, wherein the abnormality of the lymph nodes is depletion of lymph nodes relative to a wild-type mouse.

63. The transgenic mouse of claim 48, wherein the abnormality of the lymph nodes is absence of lymph nodes.

64. The transgenic mouse of claim 48, wherein the abnormality of the lymph nodes is depletion of gut associated lymphoid tissue ratio relative to a wild-type mouse.

65. The transgenic mouse of claim 48, wherein the abnormality lymphocytes comprises lymphoid infiltrates.

66. The transgenic mouse of claim 48, wherein the abnormality lymphocytes is consistent with lymphoma.

67. The transgenic mouse of claim 66, further comprising at least one of the following symptoms of disease: poor grooming, hypoactivity, increased thoracic girth, and increased abdominal girth, hepatosplenomegaly, enlarged thymus, ascites, weakness, squinting, hunching, cachexis, lethargy, and impaired respiration.

68. The transgenic mouse of claim 48, wherein the bone marrow is pale.

69. The transgenic mouse of claim 48, wherein the abnormality of the bones is brittleness.

70. The transgenic mouse of claim 48, wherein the abnormality of the bones is attached white masses.

71. A method of producing a transgenic mouse comprising a disruption in a RORγ gene, wherein the transgenic mouse exhibits at least one of the following phenotypes: a spleen abnormality, a kidney abnormality a spleen abnormality a liver abnormality, an abnormality of the thymus, an abnormality in the lymph nodes, an abnormality in the lymphocytes, an abnormality in the bone marrow, or an abnormality in the bones, the method comprising:

(a) introducing a RORγ gene targeting construct into a cell;

(b) introducing the cell into a blastocyst;

(c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

(d) breeding the chimeric mouse to produce the transgenic mouse comprising a disruption in a RORγ gene.

72. A cell derived from the transgenic mouse of claim 48 or claim 71.

73. A method of identifying an agent that ameliorates a phenotype associated with a disruption in a RORγ gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a RORγ gene; and

(b) determining whether the agent ameliorates at least one of the following phenotypes: a spleen abnormality, a kidney abnormality a spleen abnormality a liver abnormality, an abnormality of the thymus, an abnormality in the lymph nodes, an abnormality in the lymphocytes, an abnormality in the bone marrow, or an abnormality in the bones.

74. A method of identifying an agent that ameliorates a phenotype associated with a disruption in a RORγ gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a RORγ gene; and

(b) determining whether the agent ameliorates at least one of the following phenotypes: elevated serum alanine aminotransferase, elevated serum alkaline phosphatases, elevated serum aspartate aminotransferase, elevated blood urea nitrogen, and elevated blood phosphorus.

75. A method of identifying an agent that ameliorates a phenotype associated with a disruption in a RORγ gene, the method comprising:

(a) administering an agent to the transgenic mouse comprising a disruption in a RORγ gene; and

(b) determining whether the agent modulates RORγ expression in the transgenic mouse, wherein the agent has an effect on at least one of the following symptoms of disease: poor grooming, hypoactivity, increased thoracic girth, and increased abdominal girth, hepatosplenomegaly, enlarged thymus, ascites, weakness, squinting, hunching, cachexis, lethargy, and impaired respiration.

76. A method of identifying an agent that modulates RORγ gene function, the method comprising:

(a) providing a cell comprising a disruption in a RORγ gene;

(b) contacting the cell with an agent; and

(c) determining whether the agent modulates RORγ gene function, wherein the agent modulates a phenotype associated with a disruption in an RORγ gene.

77. The method of claim 45, wherein the phenotype comprises at least one of the following: a spleen abnormality, a kidney abnormality a spleen abnormality a liver abnormality, an abnormality of the thymus, an abnormality in the lymph nodes, an abnormality in the lymphocytes, an abnormality in the bone marrow, or an abnormality in the bones.

78. The method of claim 45, wherein the phenotype comprises lymphoma.

79. An agent identified by the method of claim 42, claim 43,claim 44, or claim 45.

80. A transgenic mouse comprising a disruption in an RORγ gene, wherein the transgenic mouse exhibits lymphoma.

81. An agent that modulates the activity of an RORγ gene.

82. A method of ameliorating a disease associated with a mutation of an RORγ gene, the method comprising administering to a subject in need, a therapeutically effective amount of an agent that modulates the activity of an RORγ gene.

83. A transgenic mouse comprising a disruption in BMP gene, wherein the transgenic mouse exhibits at least one of the following phenotypes: kinky tail, low body weight, or short body length.

84. A method of producing a transgenic mouse comprising a disruption in BMP gene, wherein the transgenic mouse exhibits at least one of the following phenotypes: kinky tail, low body weight, or short body length, the method comprising:

(a) introducing a BMP gene targeting construct into a cell;

(b) introducing the cell into a blastocyst;

(c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

(d) breeding the chimeric mouse to produce the transgenic mouse comprising a disruption in BMP gene.

85. A cell derived from the transgenic mouse of claim 83 or claim 84.

86. A method of identifying an agent that ameliorates a phenotype associated with a disruption in a BMP gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a BMP gene; and

(b) determining whether the agent ameliorates at least one of the following phenotypes: kinky tail, low body weight, or short body length.

87. A method of identifying an agent that modulates a BMP gene function, the method comprising:

(a) providing a cell comprising a disruption in a BMP gene;

(b) contacting the cell with an agent; and

(c) determining whether the agent modulates a BMP gene function, wherein the agent modulates a phenotype associated with a disruption in a BMP gene.

88. The method of claim 31, wherein the phenotype comprises at least one of the following: kinky tail, low body weight, or short body length.

89. An agent identified by the method of claim 86 or claim 87.

90. An agent that modulates the activity of a BMP gene.

91. A method of ameliorating a disease associated with a mutation of a BMP gene, the method comprising administering to a subject in need, a therapeutically effective amount of an agent that modulates the activity of a BMP gene.

AGGGAGAGTCTGCCCACAAGTTTTTGTATATTTTCTCACTGAGGCATCTATTCAGTTTGGGCAGCAGACA
CTGAGCAGAACGTAGCACGGCAATGCTTGGTAGCAATGCCTGTCCGGCCAGCACTCAGAAGACGGAGGCA
GGAGAATCATAGCTTCCAGTCAGCCTCTTCTACAATATAGTCAGTTGGAAGTCAGCCAGCTTAGACAACA
TGGAGAGCCTGTGCCGAAAGCCACTGGGTAAGCCCGAATCTCAGTAGCAGAGAGCTGCCCAGGGTGCGTA
CTGC:AAAAAAAAAACCTCAAACAACAGAAGTAGGGAGGTGTAAAATAAAGTGTAGGGGGGTGGAATTTA
AGCTGATGTGGACTTCCAAATAAAGTTACCTTTTAGATACCTATTTAAATCAATAGCATAGACCTGAAAC
TGTCTATCAGAAAATGTGTCTATTCTGAGGAAGGAGTGCTAACGAGGTTCTGTGAGGGGGGCCTCTGGCT
TTGAGAGGGTGTACCATCACATAAGACTCCTAAAAGCACATACTTTTATAAATTCACCATGAGCTTTAAC
ATCTTCTTTGTCATTTCGCAGACTGAGCCATGGAGTCTTTCGATGCTGACACCAATTCAACTGACCTACA
CTCACGGCCTCTGTTTCAACCCCAAGACATTGCCTCCATGGTCATTCTTGGTCTCACTTGTCTATTGGGA
CTGCTAGGCAATGGGCTGGTGCTGTGGGTAGCTGGCGTAAAGATGAAGACGACCGTGAACACAGTCTGGT
TCCTCCATCTCACCCTGGCCGATTTCCTCTGCTGCCTCTCCTTGCCCTTCTCCTTGGCTCACCTGATTCT
CCAAGGACACTGGCCCTATGGCTTGTTCCTGTGCAAACTTATCCCATCCATCATTATTCTCAACATGTTT
GCCAGTGTCTTCCTGCTTACTGCCATTAGCCTGGACCGATGTCTGATAGTACATAAGCCAATCTGGTGCC
AGAATCATCGAAACGTGAGAACCGCCTTCGCCATCTGTGGATGTGTCTGGGTGGTAGCCTTTGTGATGTG
TGTGCCCGTATTTGTATACCGTGATCTGTTCATTATGGACAATCGCAGTATATGTAGATATAATTTTGAT
TCCTCCAGGTCATATGATTATTGGGACTACGTGTACAAACTAAGTCTACCAGAAAGCAATTCTACTGATA
ACTCCACTGCTCAGCTAACTGGACATATGAATGACAGGTCAGCTCCTTCCTCTGTACAGGCAAGGGATTA
CTTTTGGACAGTTACCACTGCCCTCCAGTCACAGCCATTCCTAACATCTCCTGAAGACTCATTCTCTCTA
GATTCAGCAAACCAACAACCCCATTATGGTGGAAAGCCTCCTAATGTCCTCACAGCCGCCGTACCCAGCG
GGTTTCCTGTTGAAGATCGTAAATCCAATACACTGAACGCTGACGCTTTTCTCTCTGCTCACACAGAACT
TTTCCCTACTGCTTCTAGTGGTCATTTATACCCCTATGATTTCCAGGGGGATTATGTTGACCAATTCACG
TATGACAATCATGTGCCGACACCGCTGATGGCAATAACCATCACAAGGCTGGTGGTGGGCTTCCTGGTGC
CGTTTTTCATCATGGTAATTTGTTACAGCCTCATCGTCTTCAGAATGCGAAAAACCAACTTCACCAAGTC
TCGGAACAAAACCTTTCGGGTGGCTGTGGCTGTGGTCACTGTCTTTTTTATCTGCTGGACTCCATACCAT
CTTGTCGGAGTCCTGCTATTGATTACTGATCCAGAAAGTTCCTTGGGGGAAGCTGTGATGTCCTGGGACC
ACATGTCCATTGCTTTAGCATCTGCCAATAGTTGCTTCAACCCTTTCCTGTATGCCCTCTTGGGGAAAGA
CTTTAGGAAGAAAGCAAGACAGTCTATAAAGGGCATTCTGGAAGCAGCCTTCAGCGAAGAGCTCACGCAC
TCTACCAACTGTACCCAAGACAAAGCCTCTTCAAAAAGAAACAATATGAGTACAGATGTGTGAAGATGTG
GCCCTGGGAACCTAAGCAGAGTTCTCAGGTGAACAGTGATGGATGACATGTGAGCAGGACACTTTAGACA
ATTTGGCGACTCTCAGAGAAAGGTCTCTTATTGACATCAGCATCATTTGAAAACATTAAAGATGCAAAAT
TTCAAGCCCCATCCCAGATGTGTTGACTCAGAATCTCTGGCCCATGGGACCAGTGTTTTAACAGGCCTTC
TTGTTTCCATCAGTGTTAAGTTTTACCTCATTTGGCTTAGTCTATTCCCATCCCTGACTACACCATGTGC
AATGAATAACTTTTTCATCTGTTTTCAGTATTCTTTTTTTTTCCTTAGCATCATCTAAACTTCTAGTTTG
CATGGAAGGCTGCTCTTATTGTTCTGAATGGAAGATATTCATTTATTGTACAGTTTTGTGGTGGTGACAA
GTGATTTTTAAGTGGGGAAAGAGACACAGTAAGAAAAGATCTATGAAAGCAGGGAGTGTTGAGTTAGAGT
TTGACAGAACACAGTGCCAAATGCCACCCACTAAAAGCAACCTGAGATAATTCCAGTGTTCATGTGAGCA
AGTGAGCACAGATACACATAAACACTTTCCTACTCCTGGAGTGTTTTAGAAGTTGTAGCTTGGAGCTC
(SEQ ID NO:1)

MESFDADTNSTDLHSRPLFQPQDIASMVILGLTCLLGLLGNGLVLWVAGVKMKTTVNTVWFLHLT
LADFLCCLSLPFSLAHLILQGHWPYGLFLCKLIPSIIILNMFASVFLLTAISLDRCLIVHKPIWCQNHR
NVRTAFAICGCVWVVAFVMCVPVFVYRDLFIMDNRSICRYNFDSSRSYDYWDYVYKLSLPESNS
TDNSTAQLTGHMNDRSAPSSVQARDYFWTVTTALQSQPFLTSPEDSFSLDSANQQPHYGGKPP
NVLTAAVPSGFPVEDRKSNTLNADAFLSAHTELFPTASSGHLYPYDFQGDYVDQFTYDNHVPTP
LMAITITRLVVGFLVPFFIMVICYSLIVFRMRKTNFTKSRNKTFRVAVAVVTVFFICWTPYHLVGVLL
LITDPESSLGEAVMSWDHMSIALASANSCFNPFLYALLGKDFRKKARQSIKGILEAAFSEELTHST
NCTQDKASSKRNNMSTDV (SEQ ID NO:2)

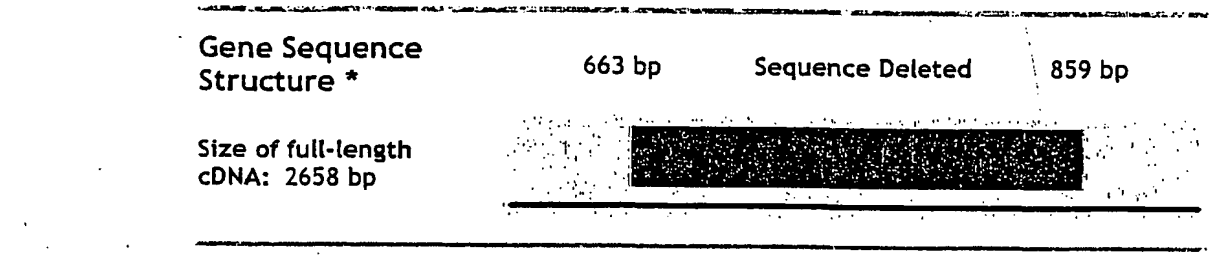
FIGURE 1

underlined = deleted in targeting construct

[ ] = sequence flanking Neo insert in targeting construct

AGGGAGAGTCTGCCCACAAGTTTTTGTATATTTTCTCACTGAGGCATCTATTCAGTTTGG
GCAGCAGACACTGAGCAGAACGTAGCACGGCAATGCTTGGTAGCAATGCCTGTCCGGCCA
GCACTCAGAAGACGGAGGCAGGAGAATCATAGCTTCCAGTCAGCCTCTTCTACAATATAG
TCAGTTGGAAGTCAGCCAGCTTAGACAACATGGAGAGCCTGT[GCCGAAAGCCACTGGGTA
AGCCCGAATCTCAGTAGCAGAGAGCTGCCCAGGGTGCGTACTGC:AAAAAAAAAACCTCA
AACAACAGAAGTAGGGAGGTGTAAAATAAAGTGTAGGGGGGTGGAATTTAAGCTGATGTG
GACTTCCAAATAAAGTTACCTTTTAGATACCTATTTAAATCAATAGCATAGACCTGAAAC
TGTCTATCAGAAAATGTGTCTATTCTGAGGAAGGAGTGCTAACGAGGTTCTGTGAGGGGG
GCCTCTGGCTTTGAGAGGGTGTACCATCACATAAGACTCCTAAAAGCACATACTTTTATA
AATTCACCATGAGCTTTAACATCTTCTTTGTCATTTCGCAGACTGAGCCATGGAGTCTTT
CGATGCTGACACCAATTCAACTGACCTACACTCACGGCCTCTGTTTCAACCCCAAGACAT
TG]CCTCCATGGTCATTCTTGGTCTCACTTGTCTATTGGGACTGCTAGGCAATGGGCTGGT
GCTGTGGGTAGCTGGCGTAAAGATGAAGACGACCGTGAACACAGTCTGGTTCCTCCATCT
CACCCTGGCCGATTTCCTCTGCTGCCTCTCCTTGCCCTTCTCCTTGGCTCACCTGATTCT
CCAAGGACACTGGCCCTAT[GGCTTGTTCCTGTGCAAACTTATCCCATCCATCATTATTCT
CAACATGTTTGCCAGTGTCTTCCTGCTTACTGCCATTAGCCTGGACCGATGTCTGATAGT
ACATAAGCCAATCTGGTGCCAGAATCATCGAAACGTGAGAACCGCCTTCGCCATCTGTGG
ATGTGTCTGGGTGGTAGCCTTTGTGATGTGTGTGCCCGTATTTGTATACCGTGATCTGTT
CATTATGGACAATCGCAGTATATGTAGATATAATTTTGATTCCTCCAGGTCATATGATTA
TTGGGACTACGTGT]ACAAACTAAGTCTACCAGAAAGCAATTCTACTGATAACTCCACTGC
TCAGCTAACTGGACATATGAATGACAGGTCAGCTCCTTCCTCTGTACAGGCAAGGGATTA
CTTTTGGACAGTTACCACTGCCCTCCAGTCACAGCCATTCCTAACATCTCCTGAAGACTC
ATTCTCTCTAGATTCAGCAAACCAACAACCCCATTATGGTGGAAAGCCTCCTAATGTCCT
CACAGCCGCCGTACCCAGCGGGTTTCCTGTTGAAGATCGTAAATCCAATACACTGAACGC
TGACGCTTTTCTCTCTGCTCACACAGAACTTTTCCCTACTGCTTCTAGTGGTCATTTATA
CCCCTATGATTTCCAGGGGGATTATGTTGACCAATTCACGTATGACAATCATGTGCCGAC
ACCGCTGATGGCAATAACCATCACAAGGCTGGTGGTGGGCTTCCTGGTGCCGTTTTTCAT
CATGGTAATTTGTTACAGCCTCATCGTCTTCAGAATGCGAAAAACCAACTTCACCAAGTC
TCGGAACAAAACCTTTCGGGTGGCTGTGGCTGTGGTCACTGTCTTTTTTATCTGCTGGAC
TCCATACCATCTTGTCGGAGTCCTGCTATTGATTACTGATCCAGAAAGTTCCTTGGGGGA
AGCTGTGATGTCCTGGGACCACATGTCCATTGCTTTAGCATCTGCCAATAGTTGCTTCAA
CCCTTTCCTGTATGCCCTCTTGGGGAAAGACTTTAGGAAGAAAGCAAGACAGTCTATAAA
GGGCATTCTGGAAGCAGCCTTCAGCGAAGAGCTCACGCACTCTACCAACTGTACCCAAGA
CAAAGCCTCTTCAAAAAGAAACAATATGAGTACAGATGTGTGAAGATGTGGCCCTGGGAA
CCTAAGCAGAGTTCTCAGGTGAACAGTGATGGATGACATGTGAGCAGGACACTTTAGACA
ATTTGGCGACTCTCAGAGAAAGGTCTCTTATTGACATCAGCATCATTTGAAAACATTAAA
GATGCAAAATTTCAAGCCCCATCCCAGATGTGTTGACTCAGAATCTCTGGCCCATGGGAC
CAGTGTTTTAACAGGCCTTCTTGTTTCCATCAGTGTTAAGTTTTACCTCATTTGGCTTAG
TCTATTCCCATCCCTGACTACACCATGTGCAATGAATAACTTTTTCATCTGTTTTCAGTA
TTCTTTTTTTTTCCTTAGCATCATCTAAACTTCTAGTTTGCATGGAAGGCTGCTCTTATT
GTTCTGAATGGAAGATATTCATTTATTGTACAGTTTTGTGGTGGTGACAAGTGATTTTTA
AGTGGGGAAAGAGACACAGTAAGAAAAGATCTATGAAAGCAGGGAGTGTTGAGTTAGAGT
TTGACAGAACACAGTGCCAAATGCCACCCACTAAAAGCAACCTGAGATAATTCCAGTGTT
CATGTGAGCAAGTGAGCACAGATACACATAAACACTTTCCTACTCCTGGAGTGTTTTAGA
AGTTGTAGCTTGGAGCTC

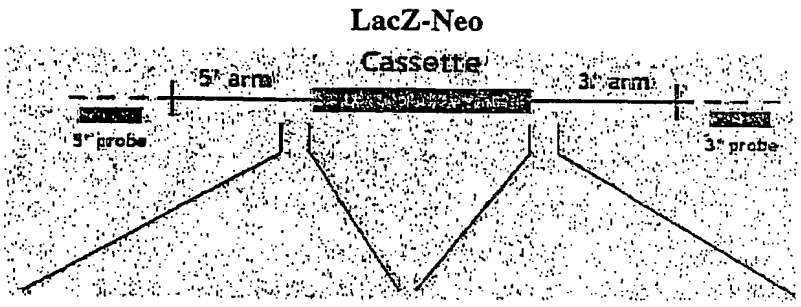
FIGURE 2A



Targeting Vector* (genomic sequence)

Construct Number: 3036

Arm Length:
5': 3.2 kb
3': 1.8 kb



5'>CGAGGTTCTGTGAGGGGGGCC TCTGGCTTTGAGAGGGTGTACCAT CACATAAGACTCCTAAAAGCACAT ACTTTTATAAATTCACCATGAGCT TTAACATCTTCTTTGTCATTTCGC AGACTGAGCCATGGAGTCTTTCGA TGCTGACACCAATTCAACTGACCT ACACTCACGGCCTCTGTTTCAACC CCAAGACATTG<3' (SEQ ID NO:3)

5'>GGCTTGTTCCTGTGCAAACTT ATCCCATCCATCATTATTCTCAAC ATGTTTGCCAGTGTCTTCCTGCTT ACTGCCATTAGCCTGGACCGATGT CTGATAGTACATAAGCCAATCTGG TGCCAGAATCATCGAAACGTGAGA ACCGCCTTCGCCATCTGTGGATGT GTCTGGGTGGTAGCCTTTGTGATG TGTGTGCCCGT<3' (SEQ ID NO:4)

Targeting Vector
Endogenous Locus

* Not drawn to scale

FIGURE 2B

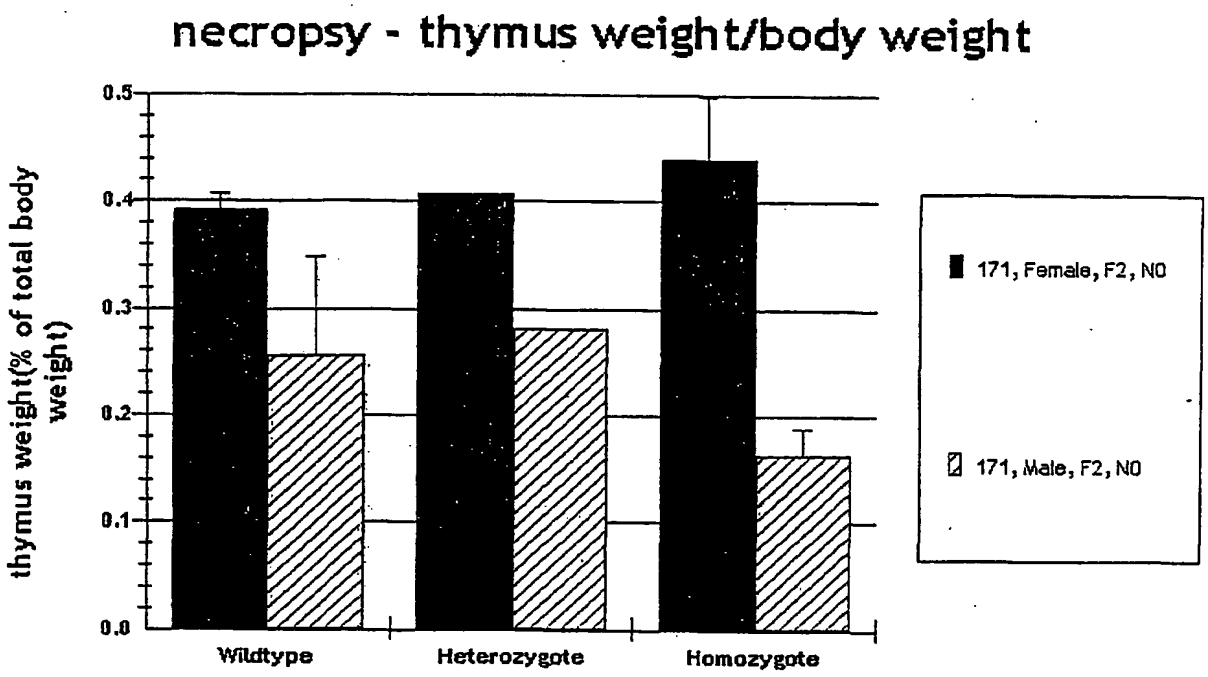


FIGURE 3

GGCACGAGGCCGTCTCCAGAAAGCAGGTATCTACGTGGCTTCCAGTCCCCAACCCCCACCCCTCGGAGCC
ACTGCCGGGAGAGGGGGGAGGTGGGCAAGGAGCAACCCTGGACCAGCGACTGTTCTGACGCACTAGCTGA
GTTCTGGGCATCCACCCTGCACTGGGCGGGGGCGACCCAAGGATGCTCTGCTGCAGGCGACCAGACAACA
GTCTCCGCCTAGGTGAGGAACAGCAAGGCATGTGATAGCAAAAGGCGGGCCCTGGCTTCTAGATTCAGCC
CCTTGAGTCCGCTTTCCATATCTCTAAGGATACCTGGGCTGTGCTGCTTGTAGCCCAGCACCCTCCTCTC
TGCTACAATTTCCTCCGGACTCTGACTGGGTGGAGACTGAGGCCAGGTTCTTGGCTCTTAGCAAAATCCT
CTCCATTGGCCATCGGTCGCAAACATCTAGATTGACTTCAGTGGGCTCGGTGGCAACACAGTCTAAACAC
AGGTGTCCTGGGACAGCAATGGATCTGCCTGTAAACTTGACCTCCTTTTCTCTCTCTACTCCCTCCTCTT
TGGAACCTAACCGCAGCTTGGACACGGAAGTCCTGCGCCCTAGTCGGCCTTTTCTCTCAGCTTTCCGAGT
GCTAGTCCTGACTTTGTTGGGCTTTCTAGCTGCGGCCACATTCACTTGGAACCTGCTGGTGCTGGCTACC
ATCCTCAAGGTACGCACCTTCCACCGAGTACCACACAACCTGGTAGCTTCCATGGCCATCTCGGATGTGC
TAGTGGCTGTGCTGGTTATGCCACTGAGCCTGGTACATGAGCTGTCTGGGCGCCGCTGGCAGCTGGGCCG
GCGTCTATGCCAGCTGTGGATCGCATGTGACGTGCTCTGCTGTACTGCCAGCATCTGGAATGTGACAGCA
ATAGCACTGGACCGCTACTGGTCAATCACGCGCCACCTGGAGTACACACTCCGTACCCGCAAGCGTGTCT
CCAATGTGATGATCCTGCTCACCTGGGCACTCTCCACTGTCATCTCTCTGGCTCCACTGCTATTTGGCTG
GGGAGAGACTTATTCTGAGCCCAGTGAGGAATGCCAAGTCAGTCGCGAGCCTTCCTACACCGTGTTCTCC
ACCGTGGGTGCCTTCTACCTGCCGCTGTGGCTGGTGCTCTTTGTGTACTGGAAAATTTACAGGGCGGCGA
AATTCCGCATGGGCTCCAGGAAGACTAACAGCGTCTCCCCCGTACCCGAAGCTGTGGAGGTGAAGAATGC
TACACAACATCCCCAGATGGTGTTCACGGCCCGCCATGCCACCGTCACCTTCCAGACAGAAGGGGATACG
TGGAGGGAGCAGAAGGAGCAAAGGGCAGCCCTCATGGTGGGCATCCTCATCGGAGTGTTTGTGCTCTGCT
GGTTCCCTTTCTTCGTCACAGAGCTCATCAGTCCCCTGTGTTCCTGGGACGTCCCTGCCATCTGGAAGAG
CATCTTCCTGTGGTTGGGCTATTCTAATTCCTTCTTCAACCCACTCATCTACACAGCATTCAACAGGAGC
TACAGCAGTGCTTTCAAGGTCTTCTTCTCCAAGCAACAATGAGAGACCACATGGGAGTGCCTTCTTCCCA
TAGCTTGTAGCTCAGTGGGTTATATTGTCCCATGAACCTTTGCAGGCTGCCCAGCTGTCTTTGAGGACAA
GATCCA (SEQ ID NO:5)

MDLPVNLTSFSLSTPSSLEPNRSLDTEVLRPSRPFLSAFRVLVLTLLGFLAAATF
TWNLLVLATILKVRTFHRVPHNLVASMAISDVLVAVLVMPLSLVHELSGRRW
QLGRRLCQLWIACDVLCCTASIWNVTAIALDRYWSITRHLEYTLRTRKRVSN
VMILLTWALSTVISLAPLLFGWGETYSEPSEECQVSREPSYTVFSTVGAFYLPL
WLVLFVYWKIYRAAKFRMGSRKTNSVSPVPEAVEVKNATQHPQMVFTARH
ATVTFQTEGDTWREQKEQRAALMVGILIGVFVLCWFPFFVTELISPLCSWDVP
AIWKSIFLWLGYSNSFFNPLIYTAFNRSYSSAFKVFFSKQQ (SEQ ID NO:6)

FIGURE 4

underlined = deleted in targeting construct

[ ] = sequence flanking Neo insert in targeting construct

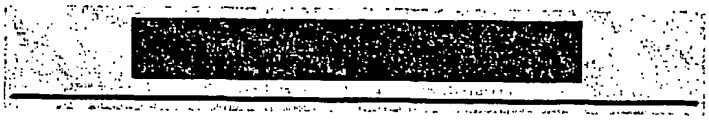
GGCACGAGGCCGTCTCCAGAAAGCAGGTATCTACGTGGCTTCCAGTCCCCAACCCCCACC
CCTCGGAGCCACTGCCGGGAGAGGGGGGAGGTGGGCAAGGAGCAACCCTGGACCAGCGAC
TGTTCTGACGCACTAGCTGAGTTCTGGGCATCCACCCTGCACTGGGCGGGGGCGACCCAA
GGATGCTCTGCTGCAGGCGACCAGACAACAGTCTCCGCCTAGGTGAGGAACAGCAAGGCA
TGTGATAGCAAAAGGCGGGCCCTGGCTTCTAGATTCAGCCCCTTGAGTCCGCTTTCCATA
TCTCTAAGGATACCTGGGCTGTGCTGCTTGTAGCCCAGCACCCTCCTCTCTGCTACAATT
TCCTCCGGACTCTGACTGGGTGGAGACTGAGGCCAGGTTCTTGGCTCTTAGCAAAATCCT
CTCCATTGGCCATCGGTCGCAAACATC[TAGATTGACTTCAGTGGGCTCGGTGGCAACACA
GTCTAAACACAGGTGTCCTGGGACAGCAATGGATCTGCCTGTAAACTTGACCTCCTTTTC
TCTCTCTACTCCCTCCTCTTTGGAACCTAACCGCAGCTTGGACACGGAAGTCCTGCGCCC
TAGTCGGCCTTTTCTCTCAGCTTTCCGAGTGCTAGTCCTGACTTTGTTGG]GCTTTCTAGC
TGCGGCCACATTCACTTGGAACCTGCTGGTGCTGGCTACCATCCTCAAGGTACGCACCTT
CCACCGAGTACCACACAACCTGGTAGCTTCCATGGCCATCTCGGATGTGCTAGTGGCTGT
GCTGGTTATGCCACTGAGCCTGGTACATGAGCTGTCTGGGCGCCGCTGGCAGCTGGGCCG
GCGTCTATGCCAGCTGTGGATCGCATGTGACGTGCTCTGCTGTACTGCCAGCATCTGGAA
TGTGACAGCAATAGCACTGGACCGCTACTGGTCAATCACGCGCCACCTGGAGTACACACT
CCGTACCCGCAAGCGTGTCTCCAATGTGATGATCCTGCTCACCTGGGCACTCTCCACTGT
CA[TCTCTCTGGCTCCACTGCTATTTGGCTGGGGAGAGACTTATTCTGAGCCCAGTGAGGA
ATGCCAAGTCAGTCGCGAGCCTTCCTACACCGTGTTCTCCACCGTGGGTGCCTTCTACCT
GCCGCTGTGGCTGGTGCTCTTTGTGTACTGGAAAATTTACAGGGCGGCGAAATTCCGCAT
GGGCTCCAGGAAGACTAACAGCGTCTCCCCCGTACCCGAAGCTGTGGAG]GTGAAGAATGC
TACACAACATCCCCAGATGGTGTTCACGGCCCGCCATGCCACCGTCACCTTCCAGACAGA
AGGGGATACGTGGAGGGAGCAGAAGGAGCAAAGGGCAGCCCTCATGGTGGGCATCCTCAT
CGGAGTGTTTGTGCTCTGCTGGTTCCCTTTCTTCGTCACAGAGCTCATCAGTCCCCTGTG
TTCCTGGGACGTCCCTGCCATCTGGAAGAGCATCTTCCTGTGGTTGGGCTATTCTAATTC
CTTCTTCAACCCACTCATCTACACAGCATTCAACAGGAGCTACAGCAGTGCTTTCAAGGT
CTTCTTCTCCAAGCAACAATGAGAGACCACATGGGAGTGCCTTCTTCCCATAGCTTGTAG
CTCAGTGGGTTATATTGTCCCATGAACCTTTGCAGGCTGCCCAGCTGTCTTTGAGGACAA
GATCCA

FIGURE 5A

Gene Sequence Structure *

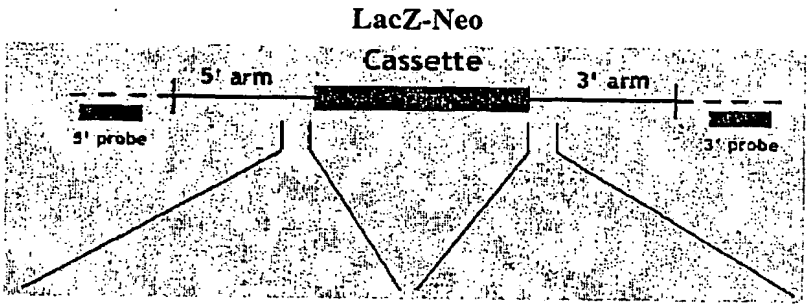
651 bp Sequence Deleted 1022 bp

Size of full-length cDNA: 1686 bp



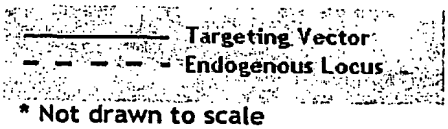
Targeting Vector* (genomic sequence)

Construct Number: 3271



Arm Length:
5': 3.5 kb
3': 2 kb

5'>GATTGACTTCAGTGGGCTCGG TGGCAACACAGTCTAAACACAGGT GTCCTGGGACAGCAATGGATCTGC CTGTAAACTTGACCTCCTTTTCTC TCTCTACTCCCTCCTCTTTGGAAC CTAACCGCAGCTTGGACACGGAAG TCCTGCGCCCTAGTCGGCCTTTTC TCTCAGCTTTCCGAGTGCTAGTCT GACTTTGTTGG<3' (SEQ ID NO:7)	5'>TCTCTCTGGCTCCACTTATTT GGCTGGGGAGAGACTTATTCTGAG CCCAGTGAGGAATGCCAAGTCAGT CGCGAGCCTTCCTACACCGTGTTC TCCACCGTGGGTGCCTTCTACCTG CCGCTGTGCGTGGTGCTCTTTGTG TACTGGAAAATTTACAGGGCGGCG AAATTCCGCATGGGCTCCAGGAAG ACCAACAGCGC<3' (SEQ ID NO:8)



* Not drawn to scale

FIGURE 5B

ATGCCGAGCCTCCCGGCCCCGCCGGCCCCGCGGCTGCTCCTCGGGCTGCTGCTGCTCGGCTCGCGGCCGG
CCAGTGGCACTGGCCCTGAGCCCCCAGCACTGCCCATCCGTTCCGAGAAGGAGCCGCTGCCTGTCCGGGG
AGCGGCAGGCTGCTCCTTCGGCGGGAAGGTCTATGCCTTGGACGAGACGTGGCACCCGGACCTGGGGGAG
CCTTTTGGGGTGATGCGCTGCGTGCTGTGCGCCTGTGAAGCGCCTCAGTGGGCTCGCCGTGGGAGGGGTC
CTGGCAGGGTCAGCTGCAAGAACATCAAACCTCAGTGCCCCACCCTGGCCTGCAGGCAGCCGCGCCAGCT
GCCAGGACACTGCTGCCAGACCTGCCCGCAGGAGCGCAGCAATCTAGACCCACAGCCCGCTGGCCTGGTC
TTCGAGTATCCAAGGGATCCAGAGCATCGCAGTTACAGTGATCGAGGGGAACCCGGCGTTGGGGAGCGGA
CCCGTGCTGATGGCCACACAGACTTTGTGGCGCTGCTGACAGGACCGAGGTCGCAGGCGGTAGCTCGTGC
TCGAGTCTCTTTGCTGCGCTCAAGTTTACGCTTCTCTGTCTCCTACCAGCGGCTGGACCGTCCCAGCAGG
GTTCGTTTCACAGATCCCACAGGCAACATCCTGTTTGAACACCCTGCAACCCCCACCCAGGATGGCCTGG
TCTGTGGGGTGTGGCGGGCAGTGCCTCGGCTGTCTGTGAGGCTCCTGAGGGCAGAGCAGCTTCGGGTAGC
CCTTGTGACATCCACTCACCCTTCAGGAGAAGTCTGGGGGCCTCTCATTTGGCAGGGTGCTCTCGCTGCA
GAGACCTTCAGTGCCATCCTGACCCTGGAAGACCCACTGCAGCGGGGTGTGGGGGGCATAGCCCTGCTCA
CCCTCAGTGACACAGAAGATTCCTTGCATTTTTTGCTGCTCTTCCGGGGTCTGCTGGGAGGACTAGCTCA
GGCTCCCTTGAAGCTTCAGATTCTCCACCAGGGACAGCTACTTCGGGAGCTCCAGGCCAACACCTCCGCT
CAGGAGCCAGGTTTTGCTGAGGTGCTGCCCAGCCTTACAGACCAAGAGATGGACTGGTTGGAGCTGGGGG
AGCTGCAGATGGTCCTAGAGAAGGCGGGTGGACCAGAGCTACGCATCAGTGGATACATCACCACCAGGCA
GAGCTGTGATGTCCTTCAAAGTGTCCTTTGTGGGGCGGATGCCCTGATCCCAGTCCAGACGGGTGCTGCT
GGGTCAGCCAGCTTCATATTGCTAGGAAATGGCTCCCTTATCTATCAGGTGCAAGTGGTAGGTACAGGTA
GCGAGGTGGTGGCCATGACACTGGAGACCAAGCCTCAGCGGAAGAACCAGCGCACTGTCCTGTGCCACAT
GGCTGGACTCCAGCCGGGAGGACACATGGCTGTGGGTATGTGCTCTGGGCTGGGTGCCCGAGGGGCTCAT
ATGCTGCTACAGAACGAGCTGTTCCTGAATGTTGGTACCAAGGACTTCCCAGATGGAGAGCTTCGGGGGC
ATGTGACTGCCCTATGCTACAGTGGGCACAGTGCCCGCTATGACAGATTGCCTGTGCCTCTGGCAGGGGC
ACTAGTGCTGCCCCCTGTGCGGAGTCAGGCAGCAGGGCACGCCTGGCTCTCCTTGGATACACATTGTCAC
TTACACTATGAAGTTCTATTGGCTGGGCTTGGTGGCTCAGAGCAAGGAACTGTCACTGCCCACCTCCTTG
GGCCTCCTGGGATGCCAGGGCCCCAGCGGCTGCTGAAGGGATTCTATGGCTCAGAGGCTCAAGGCGTGGT
AAAAGATCTGGAGCCTGTGCTGCTGCGGCACCTGGCACAGGGAACTGCCTCCCTGCTCATCACCACCAAG
AGTAGCCCCAGAGGAGAACTACGTGGGCAGGTGCACATTGCCAGTCAGTGTGAGGCGGGAGGCCTGCGTC
TGGCCTCAGAAGGAGTGCAGATGCCCTTGGCTCCAAATGGAGAGGCGGCTACGTCACCCATGCTGCCTGC
TGGCCCTGGCCCTGAAGCCCCAGTACCAGCCAAACATGGCAGCCCTGGGAGACCCCGAGATCCTAACACA
TGTTTCTTCGAGGGGCAGCAGCGGCCCCACGGGGCTCGCTGGGCACCCAACTATGACCCACTCTGCTCCC
TCTGCATCTGCCAGAGACGAACAGTGATCTGTGACCCTGTAGTATGCCCACCACCAAGCTGTCCCCACCC
GGTGCAGGCACTGGACCAGTGCTGTCCCGTGTGTCCAGAGAAACAACGCAGTAGAGACCTCCCCAGCCTA
CCAAATCTGGAGCCAGGAGAAGGCTGCTATTTTGATGGTGACCGGAGCTGGAGGGCAGCGGGTACCCGAT
GGCACCCTGTTGTGCCCCCCTTTGGCCTAATTAAGTGTGCTGTCTGCACCTGCAAGGGGGCCACGGGAGA
GGTGCACTGTGAGAAGGTGCAGTGTCCTCGCCTGGCCTGTGCTCAGCCTGTCCGTGCCAACCCCACCGAC
TGCTGCAAACAGTGTCCAGTAGGGTCAGGGACTAATGCCAAGCTGGGAGACCCCATGCAGGCTGATGGGC
CTCGGGGGTGTCGCTTTGCTGGGCAGTGGTTCCCAGAGAATCAGAGCTGGCACCCATCAGTGCCCCCCTT
TGGGGAGATGAGCTGTATTACCTGCAGATGTGGGGCAGGGGTACCCCACTGTGAGCGGGACGATTGTTCA
CCGCCGCTATCCTGTGGCTCAGGGAAGGAGAGTCGATGCTGTTCCCACTGCACAGCCCAAAGATCTTCTG
AGACTAGAACCCTTCCAGAGCTGGAGAAAGAAGCGGAGCACTCCTAGGGAGCATCCAGAAGGCCATATGA
CCAAGAGGATGGGCCTGAGCTGGGAGAAGGAGTGGTGCTGAGGACCCCGACTCTCCTGTGGGAACCCAGT
GCCTTTCGTGCCTCTGCTCTGCTTCTTCTCCTCCCCCACTACCTCTGGGAACCACAAGTCTACAAGGGGA
AAGACTGCTGGGTCAGACCAAGGCCATATCCATGCGAACTCCTGCCCTGTTTTCTTTTGGCCTCTGTCCC
ACAAGCTTAACTCTTTTCTTCTGTACAGAACACCACTGGCTTATTGGGATTCTTTAATTTATCCTCACTC
AGCACCAAGGGTTCCCTCACACCATTCCTGCTACCCCTGAGCTGAGCAGAGTCATTATTAGAGATTTTTT
TTTGTATTTATTAAAACATCTTTTTTTCATCCCA (SEQ ID NO:1)

MPSLPAPPAPRLLLGLLLLGSRPASGTGPEPPALPIRSEKEPLPVRGAAGCSFGGKVYALDETWHPDLGEPFGVMRCVLC
ACEAPQWARRGRGPGRVSCKNIKPQCPTLACRQPRQLPGHCCQTCPQERSNLDPQPAGLVFEYPRDPEHRSYSDRGEPGV
GERTRADGHTDFVALLTGPRSQAVARARVSLLRSSLRFSVSYQRLDRPSRVRFTDPTGNILFEHPATPTQDGLVCGVWRA
VPRLSVRLLRAEQLRVALVTSTHPSGEVWGPLIWQGALAAETFSAILTLEDPLQRGVGGIALLTLSDTEDSLHFLLLFRG
LLGGLAQAPLKLQILHQGQLLRELQANTSAQEPGFAEVLPSLTDQEMDWLELGELQMVLEKAGGPELRISGYITTRQSCD
VLQSVLCGADALIPVQTGAAGSASFILLGNGSLIYQVQVVGTGSEVVAMTLETKPQRKNQRTVLCHMAGLQPGGHMAVGM
CSGLGARGAHMLLQNELFLNVGTKDFPDGELRGHVTALCYSGHSARYDRLPVPLAGALVLPPVRSQAAGHAWLSLDTHCH
LHYEVLLAGLGGSEQGTVTAHLLGPPGMPGPQRLLKGFYGSEAQGVVKDLEPVLLRHLAQGTASLLITTKSSPRGELRGQ
VHIASQCEAGGLRLASEGVQMPLAPNGEAATSPMLPAGPGPEAPVPAKHGSPGRPRDPNTCFFEGQQRPHGARWAPNYDP
LCSLCICQRRTVICDPVVCPPPSCPHPVQALDQCCPVCPEKQRSRDLPSLPNLEPGEGCYFDGDRSWRAAGTRWHPVVPP
FGLIKCAVCTCKGATGEVHCEKVQCPRLACAQPVRANPTDCCKQCPVGSGTNAKLGDPMQADGPRGCRFAGQWFPENQSW
HPSVPPFGEMSCITCRCGAGVPHCERDDCSPPLSCGSGKESRCCSHCTAQRSSETRTLPELEKEAEHS (SEQ ID
NO:2)

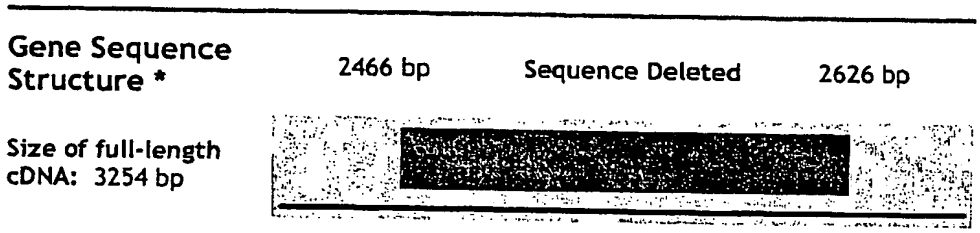
FIGURE 6

underlined = deleted in targeting construct

[ ] = sequence flanking Neo insert in targeting construct

ATGCCGAGCCTCCCGGCCCCGCCGGCCCCGCGGCTGCTCCTCGGGCTGCTGCTGCTCGGC
TCGCGGCCGGCCAGTGGCACTGGCCCTGAGCCCCCAGCACTGCCCATCCGTTCCGAGAAG
GAGCCGCTGCCTGTCCGGGGAGCGGCAGGCTGCTCCTTCGGCGGGAAGGTCTATGCCTTG
GACGAGACGTGGCACCCGGACCTGGGGGAGCCTTTTGGGGTGATGCGCTGCGTGCTGTGC
GCCTGTGAAGCGCCTCAGTGGGCTCGCCGTGGGAGGGGTCCTGGCAGGGTCAGCTGCAAG
AACATCAAACCTCAGTGCCCCACCCTGGCCTGCAGGCAGCCGCGCCAGCTGCCAGGACAC
TGCTGCCAGACCTGCCCGCAGGAGCGCAGCAATCTAGACCCACAGCCCGCTGGCCTGGTC
TTCGAGTATCCAAGGGATCCAGAGCATCGCAGTTACAGTGATCGAGGGGAACCCGGCGTT
GGGGAGCGGACCCGTGCTGATGGCCACACAGACTTTGTGGCGCTGCTGACAGGACCGAGG
TCGCAGGCGGTAGCTCGTGCTCGAGTCTCTTTGCTGCGCTCAAGTTTACGCTTCTCTGTC
TCCTACCAGCGGCTGGACCGTCCCAGCAGGGTTCGTTTCACAGATCCCACAGGCAACATC
CTGTTTGAACACCCTGCAACCCCCACCCAGGATGGCCTGGTCTGTGGGGTGTGGCGGGCA
GTGCCTCGGCTGTCTGTGAGGCTCCTGAGGGCAGAGCAGCTTCGGGTAGCCCTTGTGACA
TCCACTCACCCTTCAGGAGAAGTCTGGGGGCCTCTCATTTGGCAGGGTGCTCTCGCTGCA
GAGACCTTCAGTGCCATCCTGACCCTGGAAGACCCACTGCAGCGGGGTGTGGGGGGCATA
GCCCTGCTCACCCTCAGTGACACAGAAGATTCCTTGCATTTTTTGCTGCTCTTCCGGGGT
CTGCTGGGAGGACTAGCTCAGGCTCCCTTGAAGCTTCAGATTCTCCACCAGGGACAGCTA
CTTCGGGAGCTCCAGGCCAACACCTCCGCTCAGGAGCCAGGTTTTGCTGAGGTGCTGCCC
AGCCTTACAGACCAAGAGATGGACTGGTTGGAGCTGGGGGAGCTGCAGATGGTCCTAGAG
AAGGCGGGTGGACCAGAGCTACGCATCAGTGGATACATCACCACCAGGCAGAGCTGTGAT
GTCCTTCAAAGTGTCCTTTGTGGGGCGGATGCCCTGATCCCAGTCCAGACGGGTGCTGCT
GGGTCAGCCAGCTTCATATTGCTAGGAAATGGCTCCCTTATCTATCAGGTGCAAGTGGTA
GGTACAGGTAGCGAGGTGGTGGCCATGACACTGGAGACCAAGCCTCAGCGGAAGAACCAG
CGCACTGTCCTGTGCCACATGGCTGGACTCCAGCCGGGAGGACACATGGCTGTGGGTATG
TGCTCTGGGCTGGGTGCCCGAGGGGCTCATATGCTGCTACAGAACGAGCTGTTCCTGAAT
GTTGGTACCAAGGACTTCCCAGATGGAGAGCTTCGGGGGCATGTGACTGCCCTATGCTAC
AGTGGGCACAGTGCCCGCTATGACAGATTGCCTGTGCCTCTGGCAGGGGCACTAGTGCTG
CCCCCTGTGCGGAGTCAGGCAGCAGGGCACGCCTGGCTCTCCTTGGATACACATTGTCAC
TTACACTATGAAGTTCTATTGGCTGGGCTTGGTGGCTCAGAGCAAGGAACTGTCACTGCC
CACCTCCTTGGGCCTCCTGGGATGCCAGGGCCCCAGCGGCTGCTGAAGGGATTCTATGGC
TCAGAGGCTCAAGGCGTGGTAAAAGATCTGGAGCCTGTGCTGCTGCGGCACCTGGCACAG
GGAACTGCCTCCCTGCTCATCACCACCAAGAGTAGCCCCAGAGGAGAACTACGTGGGCAG
GTGCACATTGCCAGTCAGTGTGAGGCGGGAGGCCTGCGTCTGGCCTCAGAAGGAGTGCAG
ATGCCCTTGGCTCCAAATGGAGAGGCGGCTACGTCACCCATGCTGCCTGCTGGCCCTGGC
CCTGAAGCCCCAGTACCAGCCAAACATGGCAGCCCTGGGAGACCCCGAGATCCTAACACA
TGTTTCTTCGAGGGGCAGCAGCGGCCCCACGGGGCTCGCTGGGCACCCAACTATGACCCA
CTCTGCTCCCTCTGCATCTGCCAGAGACGAACAGTGATCTGTGACCCTGTAGTATGCCCA
CCACCAAGCTGTCCCCACCCGGTGCAGGCACTGGACCAGTGCTGTCCCGTGTGTCCAGAG
AAACAACGCAGTAGAGACCTCCCCAGCCTACCAAATCTGGAGCCAGGAGAAGGCTGCTAT
TTTGATGGTGACCGGAGCTGGAGGGCAGCGGGTACCCGATGGCACCCTGTTGTGCCCCCC
TTTGGCCTAATTAAGTGTGCTGTCTGCACCTGCAAG[GGGGCCACGGGAGAGGTGCACTGT
GAGAA]GGTGCAGTGTCCTCGCCTGGCCTGTGCTCAGCCTGTCCGTGCCAACCCCACCGAC
TGCTGCAAACAGTGTCCAGTAGGGTCAGGGACTAATGCCAAGCTGGGAGACCCCATGCAG
GCTGATGGGCCTCGGGGGTGTCGCTTTGCTGGGCAGTGGTTCCCAG[AGAATCAGAGCTGG
CACCCATCAGTGCCCCCCTTTGGGGAGATGAGCTGTATTACCTGCAGATGTGGG]GCAGGG
GTACCCCACTGTGAGCGGGACGATTGTTCACCGCCGCTATCCTGTGGCTCAGGGAAGGAG
AGTCGATGCTGTTCCCACTGCACAGCCCAAAGATCTTCTGAGACTAGAACCCTTCCAGAG
CTGGAGAAAGAAGCGGAGCACTCCTAGGGAGCATCCAGAAGGCCATATGACCAAGAGGAT
GGGCCTGAGCTGGGAGAAGGAGTGGTGCTGAGGACCCCGACTCTCCTGTGGGAACCCAGT
GCCTTTCGTGCCTCTGCTCTGCTTCTTCTCCTCCCCCACTACCTCTGGGAACCACAAGTC
TACAAGGGGAAAGACTGCTGGGTCAGACCAAGGCCATATCCATGCGAACTCCTGCCCTGT
TTTCTTTTGGCCTCTGTCCCACAAGCTTAACTCTTTTCTTCTGTACAGAACACCACTGGC
TTATTGGGATTCTTTAATTTATCCTCACTCAGCACCAAGGGTTCCCTCACACCATTCCTG
CTACCCCTGAGCTGAGCAGAGTCATTATTAGAGATTTTTTTTGTATTTATTAAAACATC
TTTTTTTCATCCCA (SEQ ID NO:9)

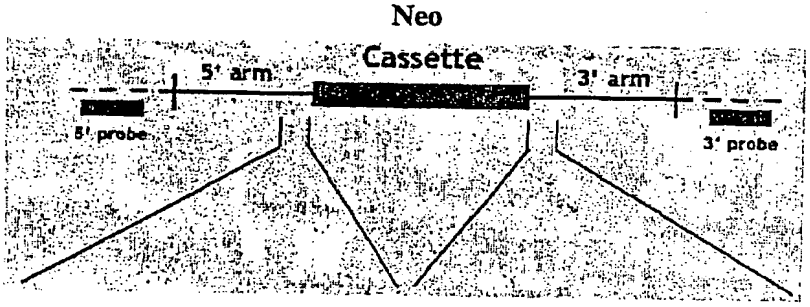
FIGURE 7A

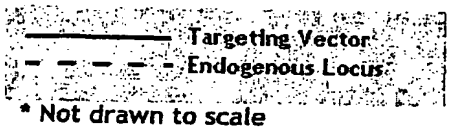


Targeting Vector* (genomic sequence)

Construct Number: 127

Arm Length:
5': 0.9 kb
3': 3.5 kb





5'>ATGCCATCCTATATACAATGA
GGACTGGATTCAGATATATAGGGC
AAACTCAGGCCTCTGAGCATTGGC
CACACATATATTTGACTTGAGGTG
GGAAAGGAGAGAGAAAGCAGTGTG
ACGGTTGTACGTCCCTCCGCCTGA
CACCCTCTCTCAATGGGTGCCTGC
CTACAGGGGGCCACGGGAGAGGTG
CACTGTGAGAA<3'(SEQ ID NO:11)

5'>AGAATCAGAGCTGGCACCCAT
CAGTGCCCCCCTTTGGGGAGATGA
GCTGTATTACCTGCAGATGTGGGG
TAAGTGAAGGACTTGTGTGATGTG
GCTAGTACTGCCTGGCCTAGAGTA
GGAAAACCTTCTCAGGGAGAGTCC
TGGAGATGGCTTTTCCTGAAGAGG
CTAAAGACTAGAGTGTCCTGTAAG
GAACCTCAGTA<3'(SEQ ID NO:12)

FIGURE 7B

GGAGGGCAGCAAGGACGGCACCAAGGGAGCTACCCCATGGACAGGGCCCCACAGAGACACCACCGGACAT
CTCGGGAGCTGCTGGCTGCAAAGAAGACCCACACCTCACAAATTGAAGTGATCCCTTGCAAGATCTGTGG
GGACAAGTCATCTGGGATCCACTACGGGGTTATCACCTGTGAGGGGTGCAAGGGCTTCTTCCGCCGCAGC
CAGCAGTGTAATGTGGCCTACTCCTGCACGCGTCAGCAGAACTGCCCCATTGACCGAACCAGCCGCAACC
GATGCCAGCATTGCCGCCTGCAGAAGTGCCTGGCTCTGGGCATGTCCCGAGATGCTGTCAAGTTTGGCCG
AATGTCCAAGAAGCAGAGGGACAGTCTACATGCAGAAGTGCAGAAACAACTGCAACAGCAGCAGCAACAG
GAACAAGTGGCCAAGACTCCTCCAGCTGGGAGCCGCGGAGCAGACACACTTACATACACTTTAGGGCTCT
CAGATGGGCAGCTACCACTGGGCGCCTCACCTGACCTACCCGAGGCCTCTGCTTGTCCCCCTGGCCTCCT
GAGAGCCTCAGGCTCTGGCCCACCATATTCCAATACCTTGGCCAAAACAGAGGTCCAGGGGGCCTCCTGC
CACCTTGAGTATAGTCCAGAACGAGGCAAAGCTGAAGGCAGAGACAGCATCTATAGCACTGACGGCCAAC
TTACTCTTGGAAGATGTGGACTTCGTTTTGAGGAAACCAGGCATCCTGAACTTGGGGAACCAGAACAGGG
TCCAGACAGCCACTGCATTCCCAGTTTCTGCAGTGCCCCAGAGGTACCATATGCCTCTCTGACAGACATA
GAGTACCTGGTACAGAATGTCTGCAAGTCCTTCCGAGAGACATGCCAGCTGCGACTGGAGGACCTTCTAC
GGCAGCGCACCAACCTCTTTTCACGGGAGGAGGTGACCAGCTACCAGAGGAAGTCAATGTGGGAGATGTG
GGAGCGCTGTGCCCACCACCTCACTGAGGCCATTCAGTATGTGGTGGAGTTTGCCAAGCGGCTTTCAGGC
TTCATGGAGCTCTGCCAGAATGACCAGATCATACTACTGACAGCAGGAGCAATGGAAGTCGTCCTAGTCA
GAATGTGCAGGGCCTACAATGCCAACAACCACACAGTCTTTTTTGAAGGCAAATACGGTGGTGTGGAGCT
GTTTCGAGCCTTGGGCTGCAGCGAGCTCATCAGCTCCATATTTGACTTTTCCCACTTCCTCAGCGCCCTG
TGTTTTTCTGAGGATGAGATTGCCCTCTACACGGCCCTGGTTCTCATCAATGCCAACCGTCCTGGGCTCC
AAGAGAAGAGGAGAGTGGAACATCTGCAATACAATTTGGAACTGGCTTTCCATCATCATCTCTGCAAGAC
TCATCGACAAGGCCTCCTAGCCAAGCTGCCACCCAAAGGAAAACTCCGGAGCCTGTGCAGCCAACATGTG
GAAAAGCTGCAGATCTTCCAGCACCTCCACCCCATCGTGGTCCAAGCCGCCTTCCCNCCACTCTATAAGG
AACTCTTCAGCACTGATGTTGAATCCCCTGAGGGGCTGTCAAAGTGATCTGGAGGAAGGACAACTTTCTA
TTTCCTTCAGCCCTCTGACCCGTCTCCCTGGACTCCCTTCACCCAGCCTTTCCCTTTCTGCACTCTATGA
AGGGTGGTATCCCTAGGAGTAAGCAAATCCTAAGACTGATTTTCTGCCCCTAGGCTTGCCTTGTAGGACA
ACAGCAGCAAGTGATGGAGAAAAGGCTTGTTATGTTTGATTTCCCATAAGTTCCACCCTGGCTTCTGGAA
GCTGTGGGGTAGATGGGATAGAGATAGGATGACCAAGTCAAATAAAAAACAGACTGACAATCAGCAGGGA
TAAATCCAGGTACCTGGGATAAGGAGAACTCAAATCTAGGCTTGAAAGCTAATAACAGTCCTTTCAATAC
CTCATTGTATTTCCCCATGGGTCCTCCTGGGGGGACATGGATCTAGCTCAGAGACTGGTGGCAAGCCCCC
AGAAGGACCTGTATATAATAAGAATATAGATTCCTG (SEQ ID NO:1)

MDRAPQRHHRTSRELLAAKKTHTSQIEVIPCKICGDKSSGIHYGVITCEGCKGFFRRSQQCNVAYS
CTRQQNCPIDRTSRNRCQHCRLQKCLALGMSRDAVKFGRMSKKQRDSLHAEVQKQLQQQQQQE
QVAKTPPAGSRGADTLTYTLGLSDGQLPLGASPDLPEASACPPGLLRASGSGPPYSNTLAKTEVQG
ASCHLEYSPERGKAEGRDSIYSTDGQLTLGRCGLRFEETRHPELGEPEQGPDSHCIPSFCSAPEVPY
ASLTDIEYLVQNVCKSFRETCQLRLEDLLRQRTNLFSREEVTSYQRKSMWEMWERCAHHLTEAIQ
YVVEFAKRLSGFMELCQNDQIILLTAGAMEVVLVRMCRAYNANNHTVFFEGKYGGVELFRALGC
SELISSIFDFSHFLSALCFSEDEIALYTALVLINANRPGLQEKRRVEHLQYNLELAFHHHLCKTHRQG
LLAKLPPKGKLRSLCSQHVEKLQIFQHLHPIVVQAAFPPLYKELFSTDVESPEGLSK (SEQ ID NO:2)

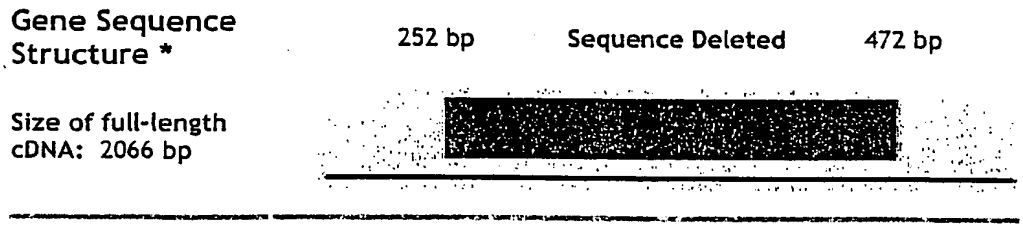
FIGURE 8

underlined = deleted in targeting construct

green = sequence flanking Neo insert in targeting construct

GGAGGGCAGCAAGGACGGCACCAAGGGAGCTACCCCATGGACAGGGCCCCACAGAGACAC
CACCGGACATCTCGGGAGCTGCTGGCTGCAAAGAAGACCCACACCTCACAAATTGAAGTG
ATCCCTTGCAAGATCTGTGGGGACAAGTCATCTGGGATCCACTACGGGGTTATCACCTGT
GAGGGGTGCAAG[GGCTTCTTCCGCCGCAGCCAGCAGTGTAATGTGGCCTACTCCTGCACG
CGTCAGCAGAA]CTGCCCCATTGACCGAACCAGCCGCAACCGATGCCAGCATTGCCGCCTG
CAGAAGTGCCTGGCTCTGGGCATGTCCCGAGATGCTGTCAAGTTTGGCCGAATGTCCAAG
AAGCAGAGGGACAGTCTACATGCAGAAGTGCAGAAACAACTGCAACAGCAGCAGCAACAG
GAACAAGTGGCCAAGACTCCTCCAGCTGGGAGCCGCGGAGCAGACACACTTA[CATACACT
TTAGGGCTCTCAGATGGGCAGCTACCACTGGGCGCCTCACCTGACCTACCCGAGGCCTCT
GCTTGTCCCCCTGGCCTCCTGAGAGCCTCAGGCTCTGGCCCACCATATTCCAATACCTTG
GCCAAAACAGAGGTCCAGGGGGCCTCCTGCCACCTTGAGTATAGTCCAGAACGAGGCAAA
GCTGAAGGCAGAGACAGCATCTATAGCACTGACGGCCAACTTACTCTTGGAAGATGTGGA
CTTCGTTTTGAGGAAACCAGGCATCCTGAACTTGGGGAACCAGAACAGGGTCCAGACAGC
CACTGCATTCCCAGTTTCTGCAGTGCCCCAGAGGTACCATATGCCTCTCTGACAGACATA
G]AGTACCTGGTACAGAATGTCTGCAAGTCCTTCCGAGAGACATGCCAGCTGCGACTGGAG
GACCTTCTACGGCAGCGCACCAACCTCTTTTCACGGGAGGAGGTGACCAGCTACCAGAGG
AAGTCAATGTGGGAGATGTGGGAGCGCTGTGCCCACCACCTCACTGAGGCCATTCAGTAT
GTGGTGGAGTTTGCCAAGCGGCTTTCAGGCTTCATGGAGCTCTGCCAGAATGACCAGATC
ATACTACTGACAGCAGGAGCAATGGAAGTCGTCCTAGTCAGAATGTGCAGGGCCTACAAT
GCCAACAACCACACAGTCTTTTTTGAAGGCAAATACGGTGGTGTGGAGCTGTTTCGAGCC
TTGGGCTGCAGCGAGCTCATCAGCTCCATATTTGACTTTTCCCACTTCCTCAGCGCCCTG
TGTTTTTCTGAGGATGAGATTGCCCTCTACACGGCCCTGGTTCTCATCAATGCCAACCGT
CCTGGGCTCCAAGAGAAGAGGAGAGTGGAACATCTGCAATACAATTTGGAACTGGCTTTC
CATCATCATCTCTGCAAGACTCATCGACAAGGCCTCCTAGCCAAGCTGCCACCCAAAGGA
AAACTCCGGAGCCTGTGCAGCCAACATGTGGAAAAGCTGCAGATCTTCCAGCACCTCCAC
CCCATCGTGGTCCAAGCCGCCTTCCCNCCACTCTATAAGGAACTCTTCAGCACTGATGTT
GAATCCCCTGAGGGGCTGTCAAAGTGATCTGGAGGAAGGACAACTTTCTATTTCCTTCAG
CCCTCTGACCCGTCTCCCTGGACTCCCTTCACCCAGCCTTTCCCTTTCTGCACTCTATGA
AGGGTGGTATCCCTAGGAGTAAGCAAATCCTAAGACTGATTTTCTGCCCCTAGGCTTGCC
TTGTAGGACAACAGCAGCAAGTGATGGAGAAAAGGCTTGTTATGTTTGATTTCCCATAAG
TTCCACCCTGGCTTCTGGAAGCTGTGGGGTAGATGGGATAGAGATAGGATGACCAAGTCA
AATAAAAAACAGACTGACAATCAGCAGGGATAAATCCAGGTACCTGGGATAAGGAGAACT
CAAATCTAGGCTTGAAAGCTAATAACAGTCCTTTCAATACCTCATTGTATTTCCCCATGG
GTCCTCCTGGGGGGACATGGATCTAGCTCAGAGACTGGTGGCAAGCCCCCAGAAGGACCT
GTATATAATAAGAATATAGATTCCTG (SEQ ID NO:1)

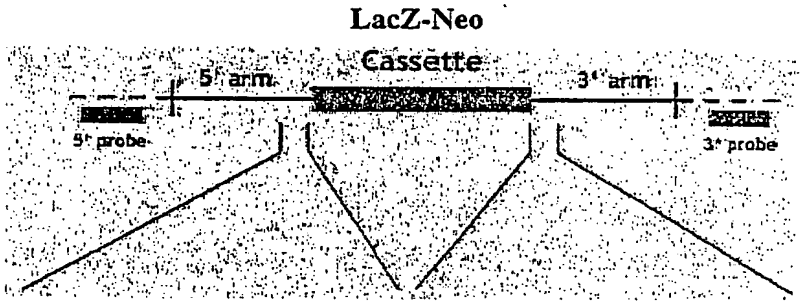
FIGURE 9A



Targeting Vector* (genomic sequence)

Construct Number: 651

Arm Length:
5': 1.6 kb
3': 2.5 kb



5'>CAGGGTCCATCACAATTATAC AGTGGAGGTTCGGGGACTTTGGTG GATGTAGAAATTCTTGAGACCAGT GCACATGAATTGGAGGTCCCTGGG ACCACCTCAAACTCCGAGAGGGTG GGATAAGCAGTTTCTGTTTCCCAG GGCTTCTTCCGCCGCAGCCAGCAG TGTAATGTGGCCTACTCCTGCACG CGTCAGCAGAA<3'(SEQ ID NO:3)	5'>CATACACTTTAGGGCTCTCAG ATGGGCAGCTACCACTGGGCGCCT CACCTGACCTACCCGAGGCCTCTG CTTGTCCCCCTGGCCTCCTGAGAG CCTCAGGCTCTGGCCCACCATATT CCAATACCTTGGCCAAAACAGAGG TCCAGGGGGCCTCCTGCCACCTTG AGTATAGTCCAGAACGAGGCAAAG CTGAAGGCAGA<3' (SEQ ID NO:4)

Targeting Vector

Endogenous Locus

* Not drawn to scale

FIGURE 9B

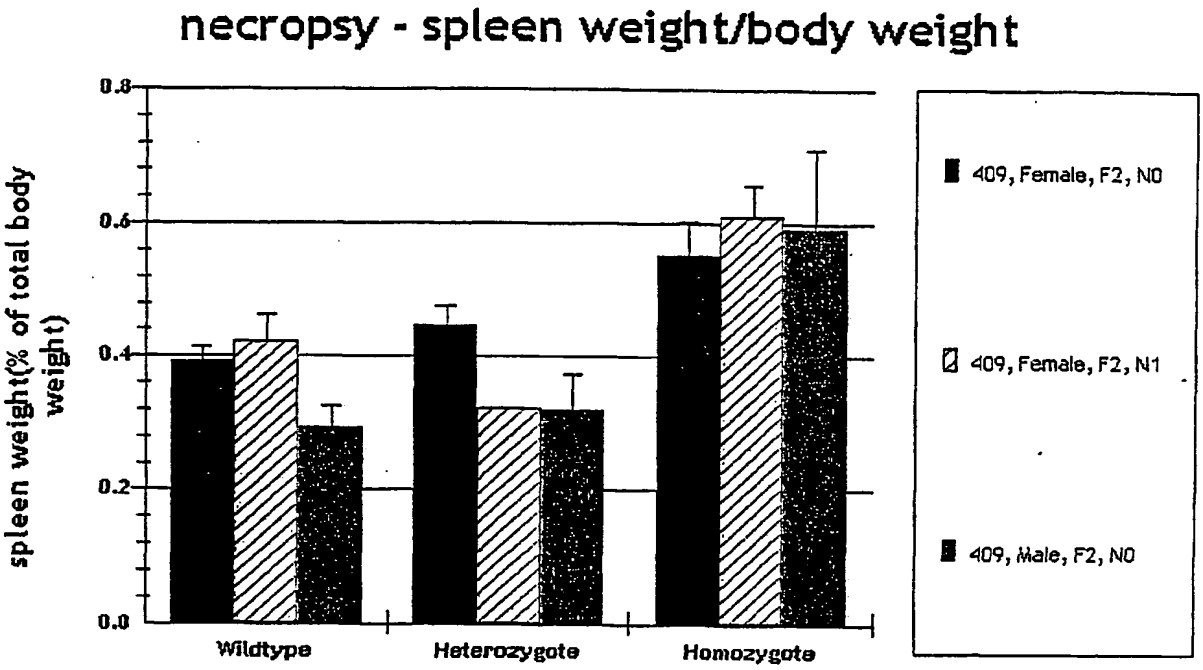

necropsy - spleen weight/body weight
spleen weight(% of total body weight)
0.8
0.6
0.4
0.2
0.0
Wildtype
Heterozygote
Homozygote
409, Female, F2, N0
409, Female, F2, N1
409, Male, F2, N0

FIGURE 10

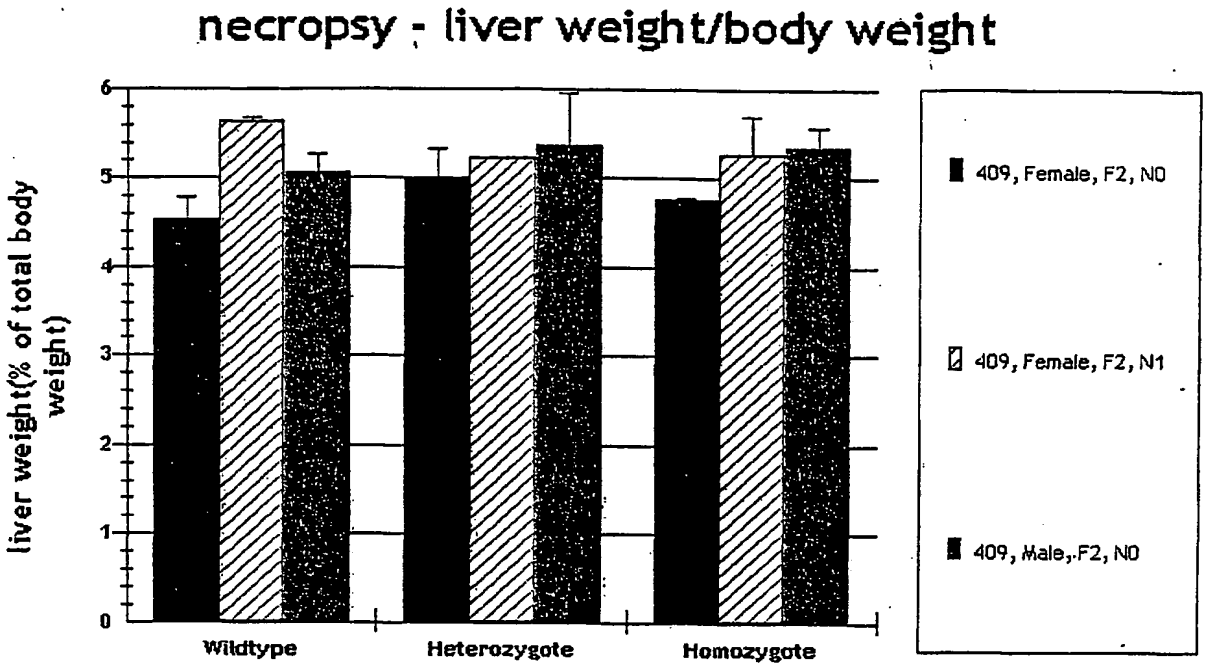

necropsy - liver weight/body weight
liver weight(% of total body weight)
0
1
2
3
4
5
6
Wildtype
Heterozygote
Homozygote
409, Female, F2, N0
409, Female, F2, N1
409, Male, F2, N0

FIGURE 11

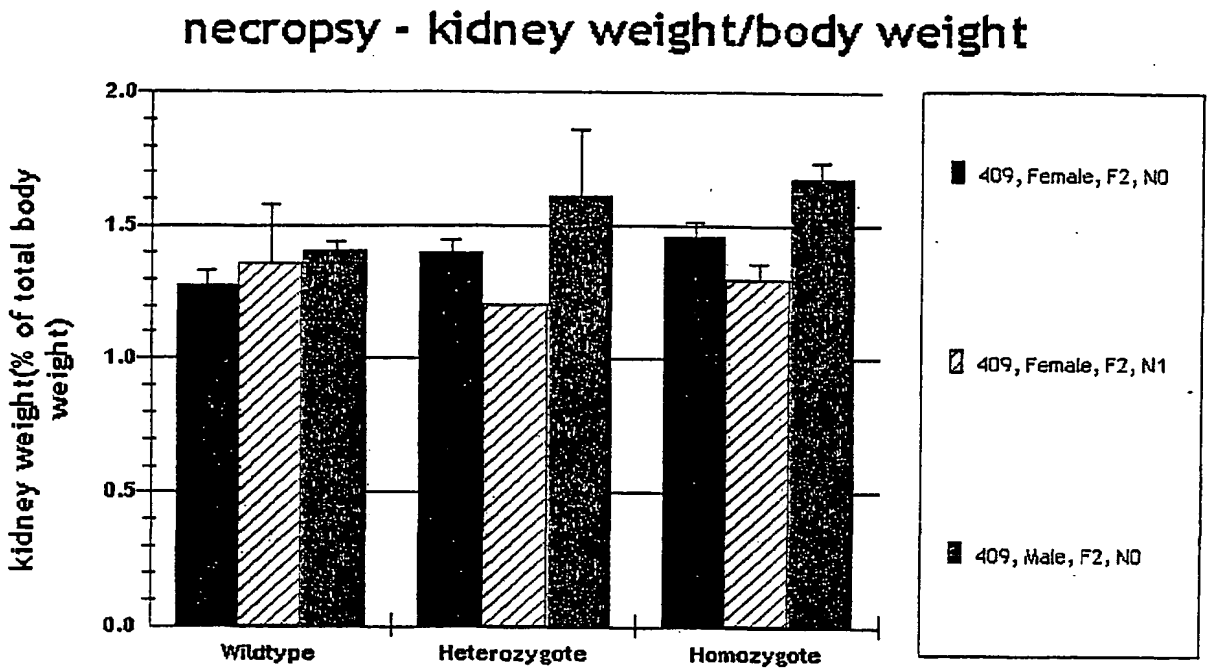

necropsy - kidney weight/body weight
kidney weight(% of total body weight)
2.0
1.5
1.0
0.5
0.0
Wildtype
Heterozygote
Homozygote
409, Female, F2, N0
409, Female, F2, N1
409, Male, F2, N0

FIGURE 12

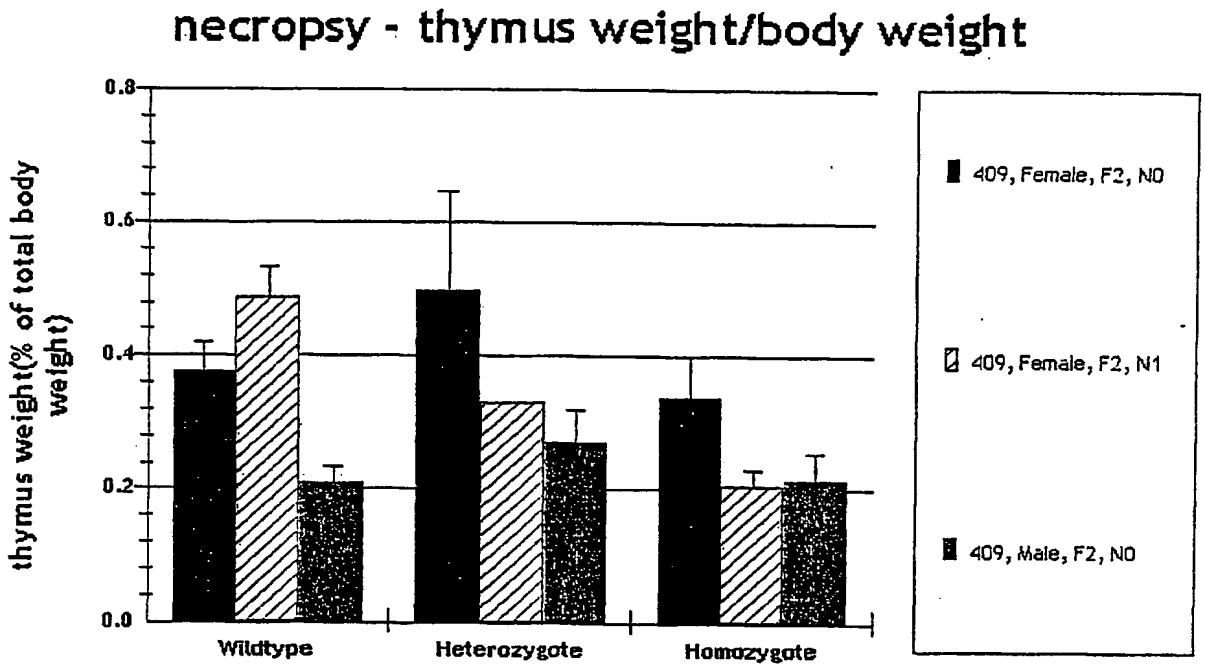

necropsy - thymus weight/body weight
thymus weight(% of total body weight)
0.8
0.6
0.4
0.2
0.0
Wildtype
Heterozygote
Homozygote
409, Female, F2, N0
409, Female, F2, N1
409, Male, F2, N0

FIGURE 13

CCACGCGTCCGCGCGGGAGCTGCTTGGAGGCTCGGCGGCCGGGAGGAGGCCGGGGCCACGCTTCTTGGAA
GCTACTGAGTGACTTCTTTGAAGAACCATGAAGTCACACTATATTGTGCTAGCTCTAGCCTCCCTGACGT
TCCTGCTGTGTCTCCCCGTGTCCCAGAGCTGTAACAAAGCACTCTGTGCCAGCGATGTGAGCAAATGCCT
CATTCAGGAGCTCTGCCAGTGCCGGCCTGGAGAAGGGAACTGCCCCTGCTGTAAGGAGTGCATGCTGTGC
CTCGGGGCCCTGTGGGACGAGTGCTGCGACTGTGTCGGTATGTGCAACCCTCGGAATTACAGCGACACCC
CGCCCACATCCAAGAGCACCGTGGAGGAGCTGCACGAGCCCATTCCGTCCCTGTTCAGGGCGCTGACGGA
GGGCGACACCCAGCTGAACTGGAACATCGTCTCCTTCCCTGTGGCAGAGGAGCTGTCACACCATGAAAAC
CTAGTCTCCTTCCTAGAAACTGTGAACCAGCTGCACCACCAAAACGTGTCTGTTCCCAGCAACAATGTCC
ACGCCCCCTTCCCCAGCGACAAAGAGCGCATGTGCACAGTGGTTTACTTTGATGACTGCATGTCCATCCA
CCAGTGTAAGATATCCTGCGAATCCATGGGTGCATCCAAGTATCGCTGGTTTCACAACGCCTGCTGCGAG
TGCATCGGTCCAGAGTGCATTGACTATGGGAGTAAAACTGTCAAGTGTATGAACTGCATGTTTTAAAGAG
GGGGAAGAAATGCAAACCAAAGCAAGAACGCTGACGATGCCGTTGACACTCCTACCTTGCAGGATGTCCC
CACAGAGTCCACTGGACCAGCTTGTACAAGGGTGAAACCTTTCTGCGTCACCTTTCTGCTTTTGCAGAGC
CATGACAGCAGCTGCCTTACAGCGTGATTTAGCAGACAGACCTGACAGGTGTGTCATGTGACAGGCAGCC
ATCAGGGTGGCCTAGCACATTTGTGGGTGCTGATGTTGTATTGCACTGTCAAAGACTGACCAGTCCTGAT
AGGGCTGGAAAGAGGAAAGGGGTGTGGGAAGGTGACTTTCTTAGCTCTCCCGTAGCAGCACACTGTCATA
GTGTCTCAGTGTCATATTGCTGTGAAGAGACACCACAACCAAGGCAGCTTATGAAGAAAGCCCTTAATTA
GGGCTCGCTTTACAGTTCCGAGGTTACTTCCTGGTCGTCGTGGTGGGGATGAGGCAGTAGGTAGCAGGCA
TGGTGCTGGAGCAGTAGCTGAATGTTGGTTCACAAACTTAGCACCTCGAGTCATGATAGAAACACCATCC
AGGCCTTGAAGCCCAACACGTAGCAACACCACTTGCCAAAGCAGGATCAAGGCTCTGTCAGCGTCCCAGG
ATGCAGTTGGGCAAGTTCTTGAGTGAAATTCCTATACATACTAACTCTGCAATTTTGCTTCTATAGTTCC
TCTTCTTGTTAACTATATGTCCAGAACCATTTCAGAATATGG
GTTTTTGTGAATAAAAAAAATAAAAAGGCCTGGTCTGCATAGATTGGGCAAGTTCCCTTGCAGCCACAGG
CTAACAGTAAAGACTTTCTATTTGGCTTTAGGAGTATCCAAGGGGTGGTCTCTCGTGAGAGTACCTCGCA
ACAGCAGCAGATGGTGTTGGAGGCTTGGCCTGCTGTGGGCTTTTGAAACCTTAAAGTCCACCCCCAGTGA
CACGCCTCCTCCAATAACACCACACCTCCTAATCCTTTCTAAGTAGTCCCTCAACTGGGAACCAAGCATT
CAGATATACGAGCCCACAAAGGCCATTCTCGTTCAAACCACCACATGTAATAAAATATATGCCACGTCAA
AAAAAAAAAAAAAAAAA (SEQ ID NO:1)

MKSHYIVLALASLTFLLCLPVSQSCNKALCASDVSKCLIQELCQCRPGEGNCP
CCKECMLCLGALWDECCDCVGMCNPRNYSDTPPTSKSTVEELHEPIPSLFRA
LTEGDTQLNWNIVSFPVAEELSHHENLVSFLETVNQLHHQNVSVPSNNVHAP
FPSDKERMCTVVYFDDCMSIHQCKISCESMGASKYRWFHNACCECIGPECID
YGSKTVKCMNCMF (SEQ ID NO:2)

FIGURE 14

red underlined = deleted in targeting construct

green = sequence flanking Neo insert in targeting construct

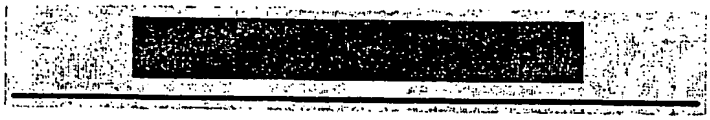
CCACGCGTCCGCGCGGGAGCTGCTTGGAGGCTCGGCGGCCGGGAGGAGGCCGGGGCCACG
CTTCTTGGAAGCTACTGAGTGACTTCTTTGAAGAACCATGAAGTCACACTATATTGTGCT
AGCTCTAGCCTCCCTGACGTTCCTGCTGTGTCTCCCCGTGTCCCAGAGCTGTAACAAAGC
ACTCTGTGCCAGCGATGTGAGCAAATGCCTCATTCAGGAGCTCTGCCAGTGCCGGCCTGG
AGAAGGGAACTGCCCCTGCTGTAAGGAGTGCATGCTGTGCCTCGGGGCCCTGTGGGACGA
GTGCTGCGACTGTGTCGGTATGTGCAACCCTCGGAATTACAGCGACACCCCGCCCACATC
CAAGAGCACCGTGGAGGAGCTGCACGAGCCCATTCCGTCCCTGTTCAGGGCGCTGACGGA
GGGCGACACCCAGCTGAACTGGAACATCGTCTCCTTCCCTGTGGCAGAGGAGCTGTCACA
CCATGAAAACCTAGTCTCCTTCCTAGAAACTGTGAACCAGCTGCACCACCAAAACGTGTC
TGTTCCCAGCAACAATGTCCACGCCCCCTTCCCCAGCGACAAAG[AGCGCATGTGCACAGT
GGTTTACTTTGATGACTGCATGTCCATCCACCAGTGTA]AGATATCCTGCGAATCCATGGG
TGCATCCAAGTATCGCTGGTTTCACAACGCCTGCTGCGAGTGCATCG[GTCCAGAGTGCAT
TGACTATGGGAGTAAAACTGTCAAGTGTATGAACTGCATGTTTTAAAGAGGGGGAAGAAA
TGCAAACCAAAGCA]AGAACGCTGACGATGCCGTTGACACTCCTACCTTGCAGGATGTCCC
CACAGAGTCCACTGGACCAGCTTGTACAAGGGTGAAACCTTTCTGCGTCACCTTTCTGCT
TTTGCAGAGCCATGACAGCAGCTGCCTTACAGCGTGATTTAGCAGACAGACCTGACAGGT
GTGTCATGTGACAGGCAGCCATCAGGGTGGCCTAGCACATTTGTGGGTGCTGATGTTGTA
TTGCACTGTCAAAGACTGACCAGTCCTGATAGGGCTGGAAAGAGGAAAGGGGTGTGGGAA
GGTGACTTTCTTAGCTCTCCCGTAGCAGCACACTGTCATAGTGTCTCAGTGTCATATTGC
TGTGAAGAGACACCACAACCAAGGCAGCTTATGAAGAAAGCCCTTAATTAGGGCTCGCTT
TACAGTTCCGAGGTTACTTCCTGGTCGTCGTGGTGGGGATGAGGCAGTAGGTAGCAGGCA
TGGTGCTGGAGCAGTAGCTGAATGTTGGTTCACAAACTTAGCACCTCGAGTCATGATAGA
AACACCATCCAGGCCTTGAAGCCCAACACGTAGCAACACCACTTGCCAAAGCAGGATCAA
GGCTCTGTCAGCGTCCCAGGATGCAGTTGGGCAAGTTCTTGAGTGAAATTCCTATACATA
CTAACTCTGCAATTTTGCTTCTATAGTTCCTCTTCTTGTTAACTATATGTCCAGAACCAT
TTCAGAATATGGGTTTTTGTGAATAAAAAAAATAAAAAGGCCTGGTCTGCATAGATTGGG
CAAGTTCCCTTGCAGCCACAGGCTAACAGTAAAGACTTTCTATTTGGCTTTAGGAGTATC
CAAGGGGTGGTCTCTCGTGAGAGTACCTCGCAACAGCAGCAGATGGTGTTGGAGGCTTGG
CCTGCTGTGGGCTTTTGAAACCTTAAAGTCCACCCCCAGTGACACGCCTCCTCCAATAAC
ACCACACCTCCTAATCCTTTCTAAGTAGTCCCTCAACTGGGAACCAAGCATTCAGATATA
CGAGCCCACAAAGGCCATTCTCGTTCAAACCACCACATGTAATAAAATATATGCCACGTC
AAAAAAAAAAAAAAAAAAA (SEQ ID NO:1)

FIGURE 15A

Gene Sequence Structure *

639 bp Sequence Deleted 707 bp

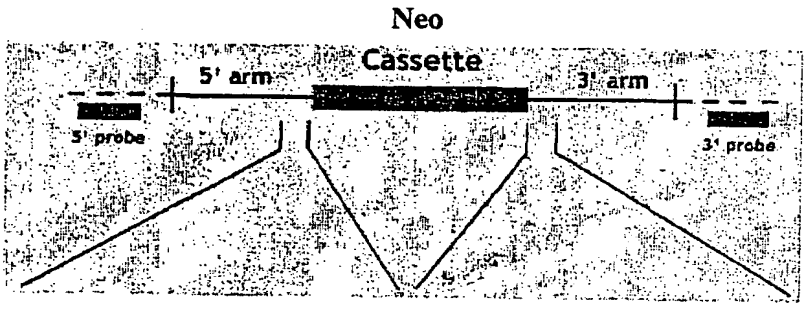
Size of full-length cDNA: 1879 bp



Targeting Vector* (genomic sequence)

Construct Number: 203

Arm Length:
5': 1.9 kb
3': 3.6 kb



Targeting Vector
Endogenous Locus
* Not drawn to scale

5'>AACAAATGAAGATCTTTTGTC
CCTCGTTTTTGTCCTGTGCTGATG
AGCAGTAAGAGGGCTAGAAAGTAA
CTGCAGGTATTCTCTGAGCAAGCG
AGCGAGTGGCCGAGCTCTTCCTGC
CTGTACTGAAATGTCCCTTTGCAT
TTCAGAGCGCATGTGCACAGTGGT
TTACTTTGATGACTGCATGTCCAT
CCACCAGTGTA<3' (SEQ ID NO:3)

5'>GTCCAGAGTGCATTGACTATG
GGAGTAAAACTGTCAAGTGTATGA
ACTGCATGTTTTAAAGAGGGGGAA
GAAATGCAAACCAAAGCAGTAAGT
CATGAAGTGTGCAGAAATCTTGGT
TCTGGTATGCTAGGAGTGTGTTAA
GTTATATGATTGTAACTGTGCTTT
TTATATCTGGTGCCTATTAGTGTA
GGTCTTTTCCA<3' (SEQ ID NO:4)

FIGURE 15B

CATCCATCAGTGTGGGGGCACGTTGATTGGTAACATGTGGGTCGTCACTGCAGCACACTGTTTTAGAACC
AATTCAAACCCTCGCCAATGGACTCTTAGTTTTGGAACAACAATAAATCCTCCCTTAATGAAAAGAGACG
TCAGAAGAATTATTATGCATGAAAGGTATCGTCCCCCAGCAAGAGACCATGACATTGCTCTAGTGCAGTT
TTCTCCCAGAGTCACCTTTTCGGATGAAGTGCGCCGAATTTGTTTGCCAGAACCCTCTGCATCTTTCCCA
CCAAATTCAACTGTCTACATCACAGGATTCGGAGCACTTTACTATGGCGGGGAATCCCAAAATGAGCTCC
GTGAAGCCAGAGTACAAATCATAAGCAATGACATCTGCAAGAAGCGACACGTGTATGGCAATGAAATAAA
ACGTGGGATGTTCTGTGCTAGATTTTCCAGTTTTTTCCACCACAGGAAATAAGTTTGTTACTTCTAAATG
CATATCAGGGAATTTAGAAGGGACGATTATATCAAAATTTGGTGGCAAAATAAAGTGAA (SEQ ID
NO:21)

FIGURE 16

underlined = deleted in targeting construct

[ ]= sequence flanking Neo insert in targeting construct

CATCCATCAGTGTGGGGGCACGTTGATTGGTAACATGTGGGTCGTCACTGCAGCACACTG
TTTTAGAAC[CAATTCAAACCCTCGCCAATGGACTCTTAGTTTTGGAAC]AACAATAAATCC
TCCCTTAATGAAAAGAGACGTCAGAAGAATTATTATGCATGAAAGGTATCGTCCCCCAGC
AAGAGACCATGACATTGCTCTAGTGCAGTTTTCTCCCAGAGTCACCTTTTCGGATGAA[GT
GCGCCGAATTTGTTTGCCAGAACCCTCTGCATCTTTCCCACCAAATTCAACTGTCTACAT
CACAGGATTCGGAGCACTTTACTATGGCG]GGGAATCCCAAAATGAGCTCCGTGAAGCCAG
AGTACAAATCATAAGCAATGACATCTGCAAGAAGCGACACGTGTATGGCAATGAAATAAA
ACGTGGGATGTTCTGTGCTAGATTTTCCAGTTTTTTCCACCACAGGAAATAAGTTTGTTA
CTTCTAAATGCATATCAGGGAATTTAGAAGGGACGATTATATCAAAATTTGGTGGCAAAA
TAAAGTGAA

FIGURE 17A

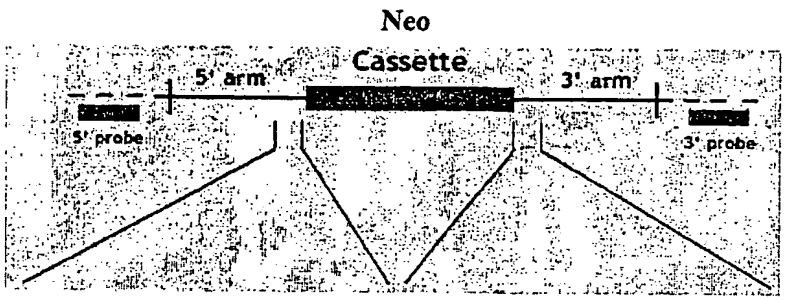
Gene Sequence Structure *

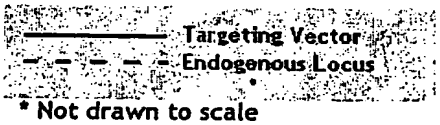
Size of EST: 549 bp

Targeting Vector* (genomic sequence)

Construct Number: 16

Arm Length:
5': 1 kb
3': 3 kb





* Not drawn to scale

5'>GCAACAGGCATATTTAAATAG CTAATGAGACTATAATTGCAACCA GCTATTGTCAGGGTGGTATGTTCT ACTGTATTAACCTAAATGCATGTT GTTAAATCATCTAAATGGTGTAGT TAATGGATACTGTACATCTCATGT ATGCTACTAATCTTTCCTAGCAAT TCAAACCCTCGCCAATGGACTCTT AGTTTTGGAAC<3'(SEQ ID NO:23)	5'>GTGCGCCGAATTTGTTTGCCA GAACCCTCTGCATCTTTCCCACCA AATTCAACTGTCTACATCACAGGA TTCGGAGCACTTTACTATGGCGGT GGGTACCTCAGAGCAGCAAGCAGA CCTCTTAGCTCTTTAACAGATCAC ATGTATTTTCTTGCCCTCCCTTTC TCCGCAGTTTTTGGTTAGTTAGTG CTCTTGTTTTA<3'(SEQ ID NO:24)

FIGURE 17B

GTGAATGTCCCCAGTCCTTTAAAAGCGCAGAGTCCGCTAGACTCTACCTCCAGTCTGTGACCTAGAGATA
AGTGAGTACAATCTAGGGCAGCCGGCGAACGTCCTGGTAGCTCTGGCTCGGATCTTCTTGAGCAGATGTC
TGGGGAGCAGACACCAATGTGTAGTATGGACCTACCTGAGGTCAAGGTGAAGACCAGCATGGCTGGCAGA
TGCCGTGTGTTCTGGTATGAGCAATACGTACAACCATGTATAGTGGAACTTGTGGGCTCCGCTCTCTTCA
TCTTCATCGGATGTCTATCGGTCATTGAGAATAGTCCGAATACTGGGCTCCTGCAGCCTGCCTTGGCTCA
TGGGCTGGCCTTGGGGCTCATCATTGCTACCTTGGGGAACATCAGCGGTGGACACTTCAACCCTGCTGTG
TCGCTGGCAGTCACAGTGATCGGAGGCCTCAAGACCATGCTGCTAATTCCCTATTGGATCTCCCAGCTAT
TTGGAGGGCTGATTGGGGCTGCCTTGGCTAAAGTGGTGAGTCCAGAGGAAAGATTCTGGAATGCATCTGG
GGCAGCCTTTGCCATCGTCCAGGAGCAGGAGCAGGTGGCAGAAGCCCTGGGGATAGAGATCATTCTGACA
ATGCTGTTGGTATTGGCTGTGTGTATGGGTGCTGTCAATGAGAAGACAATGGGCCCTTTAGCCCCATTCT
CCATTGGCTTCTCTGTCATTGTGGACATCCTGGCAGGTGGTAGCATCTCTGGAGCCTGCATGAACCCTGC
TCGTGCCTTTGGACCTGCTGTGATGGCTGGCTACTGGGACTTCCATTGGATCTACTGGCTGGGCCCCCTC
CTGGCTGGCCTCTTTGTAGGACTGCTCATTAGGCTCCTCATCGGAGATGAGAAAACCCGCCTAATTCTAA
AGTCGAGGTGAAGAAGAACTGCTGGCGACGTTCCCACTGCCTGGAGTCCTCAGCTGCTTGTCCTAAGTTG
AGGACGGAACAATTCCATTATCTCCTGCAGGACCTGAGCTAACAGGAGTCACCATGACAGCAGCTGACCT
AGGCCTGGCTTCTCAGGCAGGCAGGCAGGCAGGCGGCTGCTGCTCTGTCCTGAGCATGCTCCTGGAAACC
TGGAATTTCTCTGTGCTCATGGAGTGGGCCTGGGCCGCCTAAAGACAGGTAGAGACCGCATAGAAAAAGA
GGGTAGGCCTCTTTTCTTGCGAGAGGGCAGGGATTAGAAGGGCTGAGAAGGTGGAGAATGCACACTGTCC
ATTTCTTCAGTCCTCTTTCTTTTTTACTAGGTTGGGTTTGGAACCCAATTCATGCACACTAGGTGAGCAC
TTTGCTACTGAGCCCCTGCTCTCCACACCCTCTTTTCTCTTGATTACAGTTACAGAAGGTCTTTCAGGAG
CCCAAAGATCTTGCTTCCTCAGCTCCCAATCTGACTTTGTAATAAATG (SEQ ID NO:25)

MSGEQTPMCSMDLPEVKVKTSMAGRCRVFWYEQYVQPCIVELVGSALFIFIG
CLSVIENSPNTGLLQPALAHGLALGLIIATLGNISGGHFNPAVSLAVTVIGGLK
TMLLIPYWISQLFGGLIGAALAKVVSPEERFWNASGAAFAIVQEQEQVAEALG
IEIILTMLLVLAVCMGAVNEKTMGPLAPFSIGFSVIVDILAGGSISGACMNPAR
AFGPAVMAGYWDFHWIYWLGPLLAGLFVGLLIRLLIGDEKTRLILKSR (SEQ
ID NO:26)

FIGURE 18

underlined = deleted in targeting construct

[ ] = sequence flanking Neo insert in targeting construct

GTGAATGTCCCCAGTCCTTTAAAAGCGCAGAGTCCGCTAGACTCTACCTCCAGTCTGTGA
CCTAGAGATAAGTGAGTACAATCTAGGGCAGCCGGCGAACGTCCTGGTAGCTCTGGCTCG
GATCTTCTTGAGCAGATGTCTGGGGAGCAGACACCAATGTGTAGTATGGACCTACCTGAG
GTCAAGGTGAAGACCAGCATGGCTGGCAGATGCCGTGTGTTCTGGTATGAGCAATACGTA
CAACCATGTATAGTGGAACTTGTGGGCTCCGCTCTCTTCATCTTCATCGGATGTCTATCG
GTCATTGAGAATAGTCCGAATACTGGGCTCCTGCAGCCTGCCTTGGCTCATGGGCTGGCC
TTGGGGCTCATCATTGCTACCTTGGGGAACATCAGCGGTGGACACTTCAACCCTGCTGTG
TCGCTGGCAGTCACAGTGATCGGAGGCCTCAAGACCATGCTGCTAATTCCCTATTGGATC
TCCCAGCTATTTGGAGGGCTGATTGGGGCTGCCTTGGCTAAA[GTGGTGAGTCCAGAGGAA
AGATTCTGGAATGCATCTGGGGCAGCCTTTGCCATCGTCCAGGAGCAGGAGCAGGTGGCA
GAAGCCCTGGGGATAGAGATCATTCTGACAATGCTGTTGGTATTGGCTGTGTGTATG]GGT
GCTGTCAATGAGAAGACAATGGGCCCTTTAGC[CCCATTCTCCATTGGCTTCTCTGTCATT
GTGGACATCCTGGCAGG]TGGTAGCATCTCTGGAGCCTGCATGAACCCTGCTCGTGCCTTT
GGACCTGCTGTGATGGCTGGCTACTGGGACTTCCATTGGATCTACTGGCTGGGCCCCCTC
CTGGCTGGCCTCTTTGTAGGACTGCTCATTAGGCTCCTCATCGGAGATGAGAAAACCCGC
CTAATTCTAAAGTCGAGGTGAAGAAGAACTGCTGGCGACGTTCCCACTGCCTGGAGTCCT
CAGCTGCTTGTCCTAAGTTGAGGACGGAACAATTCCATTATCTCCTGCAGGACCTGAGCT
AACAGGAGTCACCATGACAGCAGCTGACCTAGGCCTGGCTTCTCAGGCAGGCAGGCAGGC
AGGCGGCTGCTGCTCTGTCCTGAGCATGCTCCTGGAAACCTGGAATTTCTCTGTGCTCAT
GGAGTGGGCCTGGGCCGCCTAAAGACAGGTAGAGACCGCATAGAAAAAGAGGGTAGGCCT
CTTTTCTTGCGAGAGGGCAGGGATTAGAAGGGCTGAGAAGGTGGAGAATGCACACTGTCC
ATTTCTTCAGTCCTCTTTCTTTTTTACTAGGTTGGGTTTGGAACCCAATTCATGCACACT
AGGTGAGCACTTTGCTACTGAGCCCCTGCTCTCCACACCCTCTTTTCTCTTGATTACAGT
TACAGAAGGTCTTTCAGGAGCCCAAAGATCTTGCTTCCTCAGCTCCCAATCTGACTTTGT
AATAAATG

FIGURE 19A

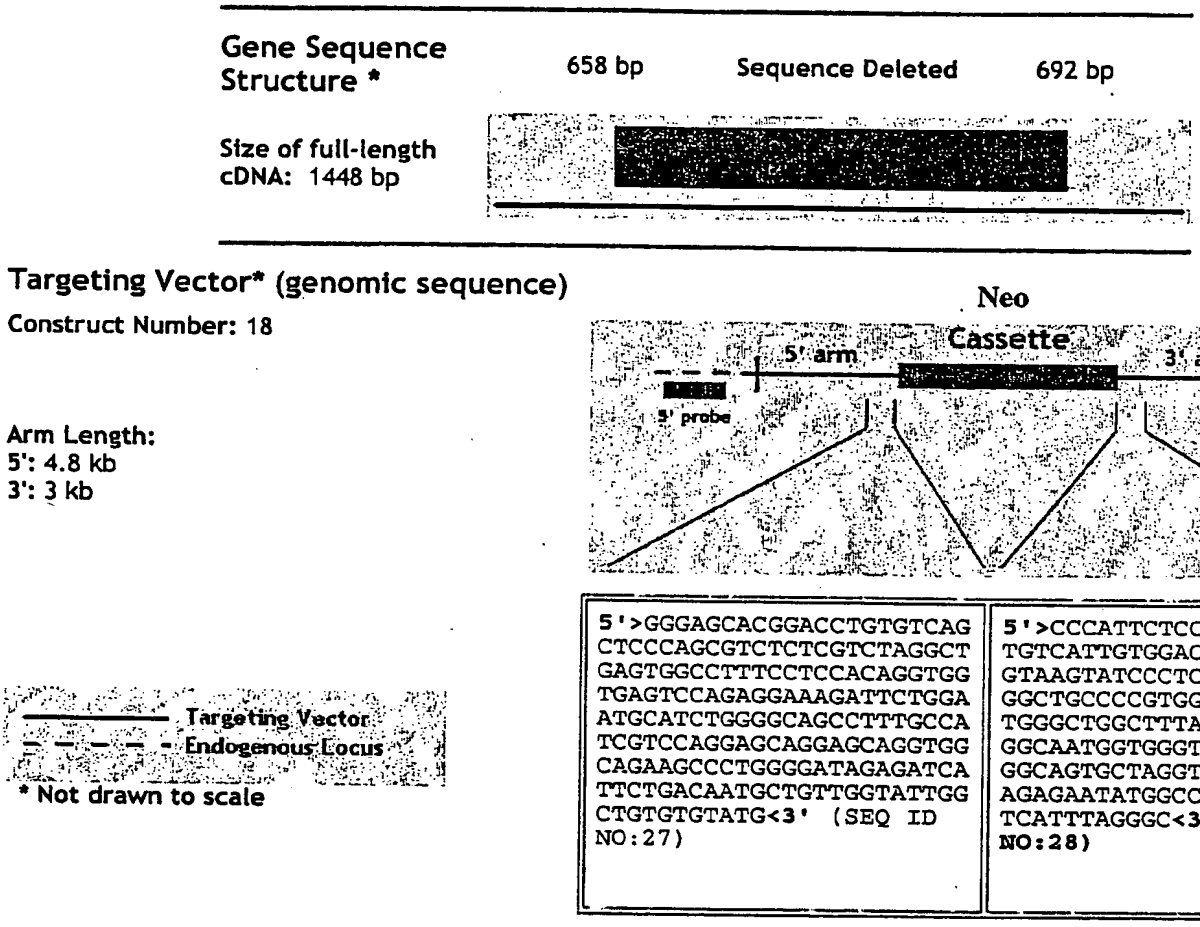

Gene Sequence Structure *
658 bp
Sequence Deleted
692 bp
Size of full-length cDNA: 1448 bp
Targeting Vector* (genomic sequence)
Construct Number: 18
Neo Cassette
5' arm
3' arm
5' probe
3' probe
Arm Length:
5': 4.8 kb
3': 3 kb
Targeting Vector
Endogenous Locus
* Not drawn to scale
5'>GGGAGCACGGACCTGTGTCAG CTCCCAGCGTCTCTCGTCTAGGCT GAGTGGCCTTTCCTCCACAGGTGG TGAGTCCAGAGGAAAGATTCTGGA ATGCATCTGGGGCAGCCTTTGCCA TCGTCCAGGAGCAGGAGCAGGTGG CAGAAGCCCTGGGGATAGAGATCA TTCTGACAATGCTGTTGGTATTGG CTGTGTGTATG<3' (SEQ ID NO:27)
5'>CCCATTCTCCATTGGCTTCTC TGTCATTGTGGACATCCTGGCAGG GTAAGTATCCCTCAAAAGTGTGAT GGCTGCCCCGTGGCAAATGGAGAG TGGGCTGGCTTTAGGCAGGCAAGA GGCAATGGTGGGTAGGTGGTATCA GGCAGTGCTAGGTGATTTTTGGCT AGAGAATATGGCCACCCTTCTACT TCATTTAGGGC<3' (SEQ ID NO:28)

FIGURE 19B

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number
WO 02/001950 A3

(51) International Patent Classification[7]: A01K 67/027, C07K 14/705, 14/51, 14/475, C12N 9/76, A61K 49/00, G01N 33/50

(21) International Application Number: PCT/US01/20795

(22) International Filing Date: 29 June 2001 (29.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/215,178	29 June 2000 (29.06.2000)	US
60/215,179	29 June 2000 (29.06.2000)	US
60/215,366	29 June 2000 (29.06.2000)	US
60/215,402	29 June 2000 (29.06.2000)	US
60/215,404	29 June 2000 (29.06.2000)	US
60/215,466	29 June 2000 (29.06.2000)	US
60/215,467	29 June 2000 (29.06.2000)	US
60/221,667	27 July 2000 (27.07.2000)	US
60/244,083	26 October 2000 (26.10.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/215,404 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,366 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,467 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,178 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,179 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/215,466 (CON)
Filed on	29 June 2000 (29.06.2000)
US	60/244,083 (CON)
Filed on	26 October 2000 (26.10.2000)
US	60/221,667 (CON)
Filed on	27 July 2000 (27.07.2000)
US	60/215,402 (CON)
Filed on	29 June 2000 (29.06.2000)

(71) Applicant *(for all designated States except US)*: **DELTAGEN, INC.** [US/US]; 740 Bay Road, Redwood City, CA 94063 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LEVITEN, Michael, W.** [US/US]; 3166 Bryant Street, Palo Alto, CA 94306 (US). **BRENNAN, Thomas, J.** [US/US]; 325 Rockwood Drive, South San Francisco, CA 94080 (US). **GUENTHER, Catherine** [US/US]; 731 Chestnut Street #109, San Carlos, CA 94070 (US). **KLEIN, Robert** [US/US]; 1044 Webster Street, Palo Alto, CA 94301 (US). **MATTHEWS, William** [GB/US]; 560 Summit Springs Road, Woodside, CA 94062 (US). **MOORE, Mark** [US/US]; 880 Blandford Avenue, Redwood City, CA 94062 (US).

(74) Agents: **LAPIZ, Mariette, A.** et al.; Deltagen, Inc., 740 Bay Road, Redwood City, CA 94063 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
27 February 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSGENIC MICE CONTAINING TARGETED GENE DISRUPTIONS

(57) Abstract: The present invention relates to transgenic animals, as well as compositions and methods relating to the characterization of gene function. Specifically, the present invention provides transgenic mice comprising mutations in a target gene. Such transgenic mice are useful as models for disease and for identifying agents that modulate gene expression and gene function, and as potential treatments for various disease states and disease conditions.

WO 02/001950 A3

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/20795

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 A01K67/027 C07K14/705 C07K14/51 C07K14/475 C12N9/76
A61K49/00 G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 A01K C07K C12N A61K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE, SEQUENCE SEARCH, CHEM ABS Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 45465 A (A & SCIENCE INVEST AB ;PEKNA MARCELA (SE)) 15 October 1998 (1998-10-15)	1-17
Y	claim 1	18-27, 31,35

Y	WO 97 28188 A (SCRIPPS RESEARCH INST ;YE RICHARD D (US)) 7 August 1997 (1997-08-07) claim 23	18-27, 31,35

X	WO 00 09129 A (BONDINELL WILLIAM E ;SMITHKLINE BEECHAM CORP (US); LEE DENNIS (US)) 24 February 2000 (2000-02-24) claims 10-13	35

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 July 2002	23. 10. 2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Lonnoy, O

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	HUMBLES ALISON A ET AL: "A role for the C3a anaphylatoxin receptor in the effector phase of asthma." NATURE (LONDON), vol. 406, no. 6799, 31 August 2000 (2000-08-31), pages 998-1001, XP002204352 ISSN: 0028-0836 figure 1 ---	1-27,31
P,X	KILDSGAARD J ET AL: "Cutting edge: targeted disruption of the C3a receptor gene demonstrates a novel protective anti-inflammatory role for C3a in endotoxin-shock." JOURNAL OF IMMUNOLOGY (BALTIMORE, MD.: 1950) UNITED STATES 15 NOV 2000, vol. 165, no. 10, 15 November 2000 (2000-11-15), pages 5406-5409, XP002204353 ISSN: 0022-1767 figure 1 ---	1-27,31
A	WO 94 26100 A (INST NAT SANTE RECH MED ;CENTRE NAT RECH SCIENT (FR); UNIV PASTEUR) 24 November 1994 (1994-11-24) ---	
A	WO 96 39511 A (INCYTE PHARMA INC) 12 December 1996 (1996-12-12) ---	
A	EP 0 814 158 A (SMITHKLINE BEECHAM CORP) 29 December 1997 (1997-12-29) ---	
A	WO 98 39659 A (ENFIELD DAVID L ;HASS G MICHAEL (US); BARD DIAGNOSTIC SCIENCES INC) 11 September 1998 (1998-09-11) ---	
A	HOLLMANN T J ET AL: "Cloning, expression, sequence determination, and chromosome localization of the mouse complement C3a anaphylatoxin receptor gene." MOLECULAR IMMUNOLOGY. ENGLAND FEB 1998, vol. 35, no. 3, February 1998 (1998-02), pages 137-148, XP002204354 ISSN: 0161-5890 ---	
P,A	HOLERS V M: "Phenotypes of complement knockouts." IMMUNOPHARMACOLOGY. NETHERLANDS AUG 2000, vol. 49, no. 1-2, August 2000 (2000-08), pages 125-131, XP002204355 ISSN: 0162-3109 -----	

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/20795

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claim 35 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [X] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

18-35 all totally 1-17 all partially

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 18-35 (all totally) and 1-17 (all partially)

   A targeting construct comprising a first polynucleotide sequence homologous to a target gene, a second polynucleotide sequence homologous to the target gene, and a selectable marker; a method of producing said targeting construct; a cell comprising a disrupted said target gene; a non-human transgenic animal comprising a disruption in said target gene; a cell derived from said transgenic animal; a method of producing a transgenic mouse comprising a disruption in said target gene; a method of identifying an agent that modulates the expression of said target gene comprising the use of said transgenic animal or of said cell; a method of identifying an agent that modulates the function of said target gene comprising the use of said transgenic animal or of said cell; Said subject-matter wherein said target gene is an anaphylatoxin C3a receptor gene.

2. Claims: 36-47 (all totally) and 1-17 (all partially)

   As for subject 1, but wherein said target gene is a chordin gene.

3. Claims: 48-82 (all totally) and 1-17 (all partially)

   As for subject 1, but wherein said target gene is a ROR-gamma gene.

4. Claims: 83-91 (all totally) and 1-17 (all partially)

   As for subject 1, but wherein said target gene is a BMP gene.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Present claims 16, 17, 33 and 34 relate to compounds defined by reference to a desirable characteristic or property, namely that it ameliorates a phenotype associated with a disruption in an anaphylatoxin C3a receptor gene, that it modulates anaphylatoxin C3a receptor gene expression, that it modulates a behavior associated with a disruption in an anaphylatoxin C3a receptor gene, or that it modulates an anaphylatoxin C3a receptor gene function. The claims cover all compounds having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for only a very limited number of such compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define a compound by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to Metrazol, as illustrated on page 49 paragraph 2 of the description section.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/20795

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9845465 A	15-10-1998	AU 7090498 A	30-10-1998
		WO 9845465 A1	15-10-1998
WO 9728188 A	07-08-1997	AU 733853 B2	31-05-2001
		AU 2256097 A	22-08-1997
		CA 2244850 A1	07-08-1997
		EP 0948536 A1	13-10-1999
		WO 9728188 A1	07-08-1997
WO 0009129 A	24-02-2000	AU 5478599 A	06-03-2000
		CA 2340053 A1	24-02-2000
		EP 1119357 A1	01-08-2001
		JP 2002522497 T	23-07-2002
		WO 0009129 A1	24-02-2000
WO 9426100 A	24-11-1994	CA 2140354 A1	24-11-1994
		EP 0659045 A1	28-06-1995
		JP 8504104 T	07-05-1996
		WO 9426100 A1	24-11-1994
		US 6030794 A	29-02-2000
		US 6031149 A	29-02-2000
WO 9639511 A	12-12-1996	AU 5972996 A	18-12-1996
		AU 718311 B2	13-04-2000
		AU 6327396 A	24-12-1996
		CA 2220530 A1	05-12-1996
		CA 2223038 A1	12-12-1996
		EP 0832282 A1	01-04-1998
		EP 0832231 A2	01-04-1998
		JP 11507513 T	06-07-1999
		JP 11506332 T	08-06-1999
		WO 9638591 A1	05-12-1996
		WO 9639511 A2	12-12-1996
EP 0814158 A	29-12-1997	EP 0814158 A2	29-12-1997
		JP 10087700 A	07-04-1998
		JP 2001228137 A	24-08-2001
		US 5942405 A	24-08-1999
WO 9839659 A	11-09-1998	AU 718842 B2	20-04-2000
		AU 2660897 A	29-10-1997
		AU 6451998 A	22-09-1998
		CA 2251344 A1	16-10-1997
		EP 0917589 A1	26-05-1999
		EP 0972200 A1	19-01-2000
		JP 2001515589 T	18-09-2001
		WO 9738136 A1	16-10-1997
		WO 9839659 A1	11-09-1998
		US 6221621 B1	24-04-2001
		US 2001023066 A1	20-09-2001